



QY	357	TGATCAAGGTGCCGAGCGACCTTGGACGAGCACCTGCCGGGCACTTCTGACTCGTTTGTA	416
Db	343	TGATTAAGGTCCCCAGCGACCTTGCAGAGCATCTGCCCGGCATTTCTGACAGCTTTTGTA	402
QY	417	GCTGGGTGCCGAGAAAGGAATGGGAGCTGCCCGGATTTCAATGATCTGAATCTGA	476
Db	403	ACTGGGTGCCGAGAAAGGAATGGGAGTTGCCCGCAGATTTCAATGATCTGAATCTGA	462
QY	477	TTGAGCAGGACCCCTGACCGTGGCCGAGAAGCTGCAGCGCACTTCTCGTCCAAATGGC	536
Db	463	TTGAGCAGGACCCCTGACCGTGGCCGAGAAGCTGCAGCGCACTTCTGACGGAATGGC	522
QY	537	GCCGCGTAGTAAGGCCCGGAGGCCCTCTTCTTTGTTCAGTTCCAGAAGGCGAGTCT	596
Db	523	GCCGCGTAGTAAGGCCCGGAGGCCCTTCTTCTTTGTGCAATTTGAGAAGGAGAGAGCT	582
QY	597	ACTTCCACCTCCATATTTCTGGTGGACAACCAACGGGGGTCAAAATCCATGCTGGGCCGT	656
Db	583	ACTTCCACATGCACGTGCTCGTGGAAACCAACGGGGGTGAATTCATGTGTTTGGGACGTT	642
QY	657	TCCTGAGTCAGATTAGGGACAAGCTGGTGCAAGACCATCTACCGCGGGATCGAGCCGACCC	716
Db	643	TCCTGAGTCAGATTTCGGAATACTGATTCAAGAAATTTACCGGGGATCGAGCCGACTT	702
QY	717	TGCCCAACTGGTTCCGCGGTGACCAAGACGCGTAATGCGCGGAGGGGGAAACAAGTGG	776
Db	703	TGCCAAACTGGTTCCGCGGTCAAAAGACCAAGAAATGCGCGGAGGGGAAACAAGTGG	762
QY	777	TGGAAGAGTGCTACATCCCCCACTAACCCTCGCCCAAGACTAGCCCGAGCTGCAGTGG	836
Db	763	TGGATGAGTGCTACATCCCCCAATTACTTGCTCCCCAAACCAAGCCTGAGCTCCAGTGG	822
QY	837	CGTGACTAATGAGAGAGTATATAAGCGCCTGTTGAACCTGGCCGAGCGCAACGCG	896
Db	823	CGTGACTAATATGGAACAAGTATTAAGCGCCTGTTGAATCTCACGGAGCGTAACGCT	882
QY	897	TCGTGGCGCAGCACCTGACCCACGTCAAGCCAGACCAAGAGCAACAAGGAATCTGA	956
Db	883	TGCTGGCGCAGCATCTGACGCACTGTCTGCAGACGCAAGAGTAGAACAAGGAATCAGA	942
QY	957	AACCCAAATTTGACGCGCCTGTCTATCCGGTCAAAAACTCCGCGCTACATGAGCTGG	1016
Db	943	ATCCCAATTTCTATGCGCGGTGATCAGATCAAAAACTTCAGCAGGTACATGAGCTGG	1002
QY	1017	TCGGGTGCTGTGGACCGGGCATCACTCCGAGAAGCAGTGGATCCAGAGAACCAAG	1076
Db	1003	TCGGGTGCTGTGGACAAGGGGATTACTCGAGAAGCAGTGGATCCAGAGAACCAAG	1062
QY	1077	CCTCGTACATCTCTTCAACGCGCTTCCAACCTCGCGGTCCAGATCAAGGCGCTTGG	1136
Db	1063	CCTCATACATCTCTTCAATGCGGCTCCAACCTCGCGGTCCCAATCAAGGCTGCTTGG	1122
QY	1137	ACAATGCCGGCAAGATCATGGCGTGAACCAATCCGCGCCCACTAACCTGTAGGCCCG	1196
Db	1123	ACAATGCCGGAAAGATTATGAGCCTGACTAAACCGCCCGCACTAACCTGTAGGCCCG	1182
QY	1197	CTCCGCCCGCGGACATTAAACCAACCGCATCTACCGCATCTTGGAGCTGAACGGCTACG	1256
Db	1183	AGCCCGTGAGGACATTTCCAGCAATCGGAATTTATAAATTTTGGAACTAAACGGGTACG	1242
QY	1257	AACTGCTACGCGGCTCCGCTTTTCTCGGCTGGGCCAGAAAGGTTCCGGAAGCGCA	1316
Db	1243	ATCCCCAATATGCGGCTTCCGCTTTCTTGGGATGGGCCACGAAAGTTCCGCAAGAGA	1302
QY	1317	ACACCATCTGGCTGTTTGGGCGGCAACCAAGGCAAGACCAATCGCGGAAGCCATCG	1376
Db	1303	ACACCATCTGGCTGTTTGGGCGCTGCACTAACGGGAAGACCAATCGCGGAGGCCATAG	1362
QY	1377	CCCAAGCCGTGCCCCCTTCTACGGCTCGCTCAACTGGAACCAATGAACCTTCCCTTCAATG	1436
Db	1363	CCCACACTGTGCCCTTCTACGGGTGCGTAACTGGAACCAATGAGAACTTCCCTTCAACG	1422
QY	1437	ATTGCGTCGACAAAGATGCTGATCTGTGGGAGGAGGCAAGATGACGGCCAAAGTCTGTG	1496

Db	1423	ACTGTGTCACAAGATGTTGATCTGTGTGGAGGAGGGGAAAGATGACCGCCAAAGTCTGTGG	1482
Qy	1497	AGTCCGCCAAGGCCATTCTCGGCGGAGCAGAAGTGCCTGTGACCAAAAAGTCAAGTCTGT	1556
Db	1483	AGTCGGCCAAAGCCATTCTCGGAGGAGCAAGTGCCTGTGACCAAGAAATGCAAGTCT	1542
Qy	1557	CCGCCAGATCGACCCCAACCCCGGTGATCGTCACTCCACACCAACATGTGCGCGTGA	1616
Db	1543	CGGCCAGATGACCCGACTCCCGTGATCGTCACTCCACACCAACATGTGCGCGTGA	1602
Qy	1617	TTGACGGGAACAGCACCACTTCGAGCACCGACGCCCTTGACAGACCCGATGTTCAAT	1676
Db	1603	TTGACGGGAACCAACGACCTTCGAAACACCGACGCCGTTGCAAGACCCGATGTTCAAT	1662
Qy	1677	TTGAACTCAACCCCGCTCTGAGCATGACTTTGGCAAGGTGACAAAGCAGAAAGTCAAG	1736
Db	1663	TTGAACTCAACCCCGCTCTGAGCATGACTTTGGGAAGTTCACCAAGCAGAAAGTCAAG	1722
Qy	1737	AGTTCTTCCGCTGGCGCAGATCACGTGACCGAGGTGGCGCATGATTCTACGTCAAG	1796
Db	1723	ACTTTTCCGCTGGCAAAAGATCACGTGTTGAGTGAAGCATGAATTTACGTCAAAA	1782
Qy	1797	AGGGTGAGCCAAACAAAGACCCCGCCGATGACCGGATAAAGCGAGCCCAAGCGG	1856
Db	1783	AGGGTGAGCCCAAGAAAGACCCCGCCGATGACCGAGATATAAGTAGCCCAAGCGG	1842
Qy	1857	CCTGCCCCCTAGTCGCGGATCCATCGACGCTACAGCGCGGAAGAGCTCCGGTGACTTG	1916
Db	1843	TGCGCGAGTCAGTTGCGCAGCCATCGACGCTACAGCGCGGA--AGCTTCGATCACTACG	1899
Qy	1917	CCGACAGGTACCAAAACAATGTTCTCGTCACGCGGGCATGCTTCAGATGCTGTTCCCT	1976
Db	1900	CAGACAGGTACCAAAACAATGTTCTCGTCACGCGGGCATGAATCTGATGCTGTTCCCT	1959
Qy	1977	GCAAGACATGCGAGAGATGAATCAGAATTTCAACTTGCTTCACGCACGGGACGAG	2036
Db	1960	GCAGACAATGCGAGAGATGAATCAGAATTTCAATATCTGCTTCACTCAGCAGAGAAAG	2019
Qy	2037	ACTGTTCAAGTGTCTCCCGCGGTGTCAGAAATCTCAACCGGT--CGTCAGAAAGAGA	2093
Db	2020	ACTGTTAGAGTGTCTCC--CGTCAGAAATCTCAACCGGTTCTGTCGTCAAAAGG	2076
Qy	2094	CGTATCGGAACTCTGTGCCATTCATCATCTGTGGGGCGGGCTCCCGAGATGCTTGT	2153
Db	2077	CGTATCAGAACTGTGTCTACATTCATCATATCATGGGAAAGGTGCCAGA--CGCTTGA	2133
Qy	2154	CGGCCTGCGATCTGTCTCAACGTGGAACCTGGATGACTGTGTTCTGAGCAATAATGACTT	2213
Db	2134	CTGCCTGCGATCTGTCTCAATGTGGAATTGGATGACTGCATCTTGAAACAATAATGATTT	2193
Qy	2214	AAACCAAGTATGCTGCGGATGTTATCTTCCAGATTGGCTCGAGGACAACTCTCTGAG	2273
Db	2194	AAATCAGTATGCTGCGGATGTTATCTTCCAGATTGGCTCGAGGACACTCTCTGAA	2253
Qy	2274	GGCATTGCGAGTGTGAGGACTTGAACCTGAGCCCGAAGCCCAAGCCAAAGCAGCAAA	2333
Db	2254	GGAATTAAGACAGTGTGGAAGCTCAAACTGGCCCAACCAACCAAGCCCGCAGAGCGG	2313
Qy	2334	AAGCAGGACGACGCGCGGGGTCTGTGCTTCTGAGCTACAAGTACCTCGAGCCCTCAAC	2393
Db	2314	CATAAGGACGACGAGGGGTCTGTGCTTCTGAGGTAACAAGTACCTCGAGCCCTCAAC	2373
Qy	2394	GGAATCGACAAAGGAGAGCCCGGTCAACGAGGACGACGCCCGGCTCGAGCAGACAAAG	2453
Db	2374	GGAATCGACAAAGGAGAGCCCGGTCAACGAGGACGACGCCCGGCTCGAGCAGACAAAG	2433
Qy	2454	GCCTACGACGACGCTCAAAAGCGGGTGACAAATCGTACCTGCGGTATAACACGCGGAC	2513
Db	2434	GCCTACGACGCGGACGCTCGACAGCGGAGACAAACCGTACCTCAAGTACCAACGCGGAC	2493
Qy	2514	GCCGAGTTTCAGAGCGTCTGCAAGAAGATACGTTTGGGGGCAACCTCGGGGAGCA	2573



Db 2494 GCGAGTTTCAGAGCGCCTTAAAGAAATACGTCCTTTGGGGGCAACCTCGAGCAGCA 2553  
QY 2574 GTCTTCAGGCCAAGAGCGGGTCTCGAACTCTCGGTCTGGTTGAGGAAGCGCTAG 2633  
Db 2554 GTCTTCAGGCCAAGAGCGGGTCTCGAACTCTGGGCTGGTTGAGGAACCTGTTAG 2613  
QY 2634 ACCGCTCTGGAAGAAACGTCCGGTAGAGCAGTCGCCACAAGCCAGACTCCTCTCG 2693  
Db 2614 ACCGCTCCGGGAAAAAGAGCGCGGTAGAGCACTCTCTGTGGAGCCAGACTCCTCTCG 2673  
QY 2694 GGCATCGGCAAGACAGCGCCAGCAGCCCGCTAAAGAGACTCAATTTTGTGACAGTGGC 2753  
Db 2674 GGAACCGGAAGGCGGGCCAGCAGCTTCCAAGAAAAAGATTGAATTTTGTGACAGTGG 2733  
QY 2754 GACTCAGAGTCAGTCCCGGATCCAACCTCTCGGAGAACCTTCAGCAACCCCGCTGCT 2813  
Db 2734 GACGCACTCAGTACCTGACCCCAAGCCTCTCGGACAGCAGCAGCCCTCTGCT 2793  
QY 2814 GTGGACCTACTACAAATGCTTCAGCGGGTGGCGCACCAATGGCAGACAATTAACGAAGC 2873  
Db 2794 CTGGAACTAATACGATGGCTACAGGCACTGGCGCACCAATGGCAGACAATTAACGAGGC 2853  
QY 2874 GCCGACGAGTGGGTAATGCTCAGAAATTGGCATTGGCAATTCACATGCTGGCGAC 2933  
Db 2854 GCCGACGAGTGGGTAATGCTCAGAAATTGGCATTGGCAATTCACATGATGGCGAC 2913  
QY 2934 AGAGTCATCACCAACAGCAGCCCGCACTGGGGCTTGGCCCACTACAATAACCACTCTAC 2993  
Db 2914 AGAGTCATCACCAACAGCAGCCCGCACTGGGGCTTGGCCCACTACAATAACCACTCTAC 2973  
QY 2994 AAGCAATCTCAGTGTTCACGGGGGGCCAGCAACGACAACCACTACTTGGGCTACAGC 3053  
Db 2974 AAACAATTTCCAGCCCAATCA--GAGGCTCGAAGCAATCACTATTGGCTACAGC 3030  
QY 3054 ACCCCCTGGGGTATTTGATTTCAACAGATTCCACTGCCACTTTTCAACCACTGACTGG 3113  
Db 3031 ACCCTTGGGGTATTTGATTTCAACAGATTCCACTGCCACTTTTCAACCACTGACTGG 3090  
QY 3114 CAGGCACTCATCAACAACAATGGGGAATTCGGGCCCAAGACTCAACTTCAAACTCTTC 3173  
Db 3091 CAAGACTCATCAACAACAATGGGGAATTCGGGCCCAAGACTCAACTTCAAGCTCTTT 3150  
QY 3174 AACATCCAAGTCAAGAGGTCAAGCAAGATGATGCGCTCAACAACCATGCTAATACTT 3233  
Db 3151 AACATCCAAGTCAAGAGGTCAAGCAAGATGATGCGCTCAACAACCATGCTAATACTT 3210  
QY 3234 ACCAGCAGGTTCAGTCTTCTCGGACTCGGAGTACCAAGCTTCCGTCCGCTCT 3293  
Db 3211 ACCAGCAGGTTCAGTCTTCTCGGACTCGGAGTACCAAGCTTCCGTCCGCTCT 3270  
QY 3294 GCGCAGCAGGGCTGCTCCCTCGTTCGGGGGAGCGTGTCAATGATTCGCAATACGGC 3353  
Db 3271 GCGCAGCAGGGCTGCTCCCGCTTCCAGCAGACGTCTTCATGCTGCACAGTATGGA 3330  
QY 3354 TACCTGACGCTCAACAATGGGCAAGCCGTTGAGCGTTCATCTTTACTGCTGGA 3413  
Db 3331 TACCTGACGCTCAACAATGGGCAAGCCGTTGAGCGTTCATCTTTACTGCTGGA 3390  
QY 3414 TATTTCCCTTCTCAGATGCTGAGAACGGGCAACAACCTTTACTGCTGAG 3473  
Db 3391 TACTTCTTCTCAGATGCTGCTGACCGGAACAACCTTTACTGCTGAG 3450  
QY 3474 GAAAGTCTTTTCCACAGCAGTACCGGCAACAGCCAGAGCTTGAACCGGCTGATGAATCT 3533  
Db 3451 GACGTTCTTTTCCACAGCAGTACCGCTCAACAGCCAGAGCTTGAACCGTCTCATGAATCT 3510  
QY 3534 CTGATGACCAATATCTGTATTACTGAACAGAACTCAAAATCAGTCCGGAAGTGGCCAA 3593  
Db 3511 CTGATGACCAATATCTGTATTACTGAAGAAACAACATCCAAAGTGAACCAACAGCAG 3570  
QY 3594 AACAGGACTTGTGTTAGCCGTGGTCTCCAGCTGGCATGTCTGTTAGCCCAAAAC 3653  
Db 3571 CAGTCAAGGCTTCACTTTCTCAGGCCGGAAGCAGTGAACATTCGGGACCACTAGGAAC 3630

QY 3654 TGGCTACCTGAGCCCTGTTATCGGCAGCAGCGGCTTCTAAACAAAAACAGACAAC 3713  
Db 3631 TGGCTTCTGAGCCCTGTTATCCGCCAGCAGCGAGTATCAAGACATCTGCGGATACAA 3690  
QY 3714 AACAGCAATTTTACCTGCACTGTGCTTCAAAATATTAACCTCAATGGGCGCTGAATC 3773  
Db 3691 AACAGTAACTCTGTGACTGAGCTACCAAGTACCACTCAATGGCAGAGACTCTCTG 3750  
QY 3774 ATCAACCTTGCACTGTATGGCTCACAACAAGAGCAGAAAGCAAGTCTTTCCATG 3833  
Db 3751 GTGAATCCGGCCCGGCATGGCAAGCCACAAGAGCAGATGAAGAAAGTTTTCCTCAG 3810  
QY 3834 AGCGGTGTCATGATTTTGAAGAAAGAGAGCGCCGAGCTTCAAAACACTGATGCAAT 3893  
Db 3811 AGCGGGTCTCATCTTTGGGAAGCAAGGCTCAGAGAAACAATGTGACATGGAAG 3870  
QY 3894 GTCATGATTACAGACGAGGAATTAAGCCACTTAACCTGTGGCCACCGAAGATTT 3953  
Db 3871 GTCATGATTACAGACGAGGAATTAAGCCACTTAACCTGTGGCCACCGAAGATTT 3930  
QY 3954 GGAACCGTGGCAGTCAATTTCCAGAGCAGCAGCAGACCCCTGCGACCGGAGATGCAAT 4013  
Db 3931 GGTCTGTATCTACCAACTCCAGAGAGGCAACAGACAGCACTACCGCAGATGTAAC 3990  
QY 4014 GCTATGGAGCATTAACCTGCACTGTGTGGCAAGATAGACGTGTACCTGCAAGTCCC 4073  
Db 3991 ACACAAGGCGTCTTCCAGGCACTGTGTGGCAGAGCAGATGTACTTACGGGCCC 4050  
QY 4074 ATTTGGCCAAATTCCTCACAACAGATGACACTTTCACCCGCTCTCTTATGGCGGC 4133  
Db 4051 ATCTGGGCAAGATTCACACACAGAGCAGACATTTCAACCCCTCTCCCTCATGGGTGA 4110  
QY 4134 TTTGACTCAAGAACCCGCTCTCAGATCCTCATCAAAAAACAGCCTGTCTCGCAAT 4193  
Db 4111 TTGGAATTAACACCCCTCTCACAAGATTTCTCATCAAGAACACCCCGTACTCGCAAT 4170  
QY 4194 CCTCCGGCGAGTTTTCAGCTCAAAAGTTGCTTCAATTCATCAACCAATATCTCCACAGA 4253  
Db 4171 CTTTCAACCACTTCACTGCGGCAAGTTGCTTCTTCAATCAACAGTACTCCAGCGGA 4230  
QY 4254 CAAGTAGTGAATTAATGGAGCTGCAGAAAGAAACAGCAAGCGCTGGAATCCC 4313  
Db 4231 CAGGTAGCTGAGATGAGTGGAGCTGCAGAAAGAAACAGCAAGCGCTGGAATCCC 4290  
QY 4314 GAAGTCACTACACATCCAATTAATGCAAAATCTGCCAAGTGAATTTACTGTGCAAC 4373  
Db 4291 GAATTCAGTACACTTCAACTACAACAAGTCTGTTAATGTGACTTACTGTGCACT 4350  
QY 4374 AATGACTTATCTGAGCTCGCCCATTTGGCACCCGTTACCTTACCCTCCCTGTA 4433  
Db 4351 AATGCGGTATTCAGAGCTCGCCCATTTGGCACCAATGCTGATCTGTAATCTGTA 4410  
QY 4434 TTACGTGTTAATCAATTAACCGGTGATTCGTTTCAATGAACTTTGGTCTCTGCTT 4493  
Db 4411 TTGCTGTTAATCAATTAACCGTTAATTCGTTTCAATGAACTTTGGTCTCTGCTT 4470  
QY 4494 CTTATCTTATCGGTTACATGCTTATAGCTTACATTAACCTGTTGGTGGCTTGGC 4553  
Db 4471 TCTTCTTATCTAGTTTTCATGCTTACGTAGATTAAGTATGAGCGGGTAAATCAATTAAC 4530  
QY 4554 ATAAAGACTTACGTATCGGGTTACCCCTAGTGAAGTGGCCACTCCCTCTGCG 4613  
Db 4531 TACAAGGA-----ACCCTAGTGAAGTGGCCACTCCCTCTGCG 4574  
QY 4614 CGCTCGCTCGCTCGGTGGGGCTGCGGACCAAGTCCGAGAGCGCAGAGCTGCTCT 4673  
Db 4575 CGCTCGCTCGCTCACTGAGGCCGGGCGACCAAGTCCGCGCGGCTTGGCCG 4634  
QY 4674 GCCGCCCCCAGCAGCGAGCGGCGCAGAGAGGAGTGGCAA 4718  
Db 4635 GCGGCGCTCAGTGAAGCGAGCGCGGCGCAGAGAGGAGTGGCAA 4679

RESULT 2

US-08-254-358-1

Sequence 1, Application US/08254358

Patent No. 5658785

GENERAL INFORMATION:

APPLICANT: Johnson, Philip R.

TITLE OF INVENTION: Adeno-Associated Virus Materials and

TITLE OF INVENTION: Methods

NUMBER OF SEQUENCES: 3

CORRESPONDENCE ADDRESS:

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun

STREET: 6300 Sears Tower, 233 S. Wacker Drive

CITY: Chicago

STATE: Illinois

COUNTRY: USA

ZIP: 60606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/254,358

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: No. 5658785and, Greta E.

REGISTRATION NUMBER: 35,302

REFERENCE/DOCKET NUMBER: 31975

TELECOMMUNICATION INFORMATION:

TELEPHONE: (312) 474-6300

TELEFAX: (312) 474-0448

TELEX: 25-3856

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 4680 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-254-358-1

Query Match 65.3%; Score 3082.4; DB 1; Length 4680;

Best Local Similarity 80.0%; Pred. No. 0;

Matches 3781; Conservative 0; Mismatches 891; Indels 54; Gaps 11;

QY 1 TTGCCACTCTCTCTCTGCGCGCTCGCTCGCTGGGGCGCTGGACCAAGGTCCGC 60  
DB 1 TTGGCCACTCTCTCTCTGCGCGCTCGCTCGCTGAGGCGCGGCGCAAAAGTCCGC 60  
QY 61 AGACGGCAGAGCTTGTCTCTGCGCGCGCGCGCGCGCGAGCGAGCGCGAGAGGAGTG 120  
DB 61 CGACGCCCGGCTTTGCCCGCGCGCGCTCAGTGAAGCGAGCGCGAGAGGAGTG 120  
QY 121 GGCACTCCATCACTAGGGGTAATCGCGAAGCGCGCTCCACGCTGCGCGCTCAGCGTGA 180  
DB 121 GGCACTCCATCACTAGGGGTAATCGCGAAGCGCGCTCCACGCTGCGCGCTCAGCGTGA 180  
QY 181 CGTAATTAATGCTATAGG--GAGTGTCTGTATTAAGTGTCAAGTGTCTTTGC 237  
DB 181 CGTAATTAATGCTATAGG--GAGTGTCTGTATTAAGTGTCAAGTGTCTTTGC 237  
QY 164 CGTAATTAATGCTATAGG--GAGTGTCTGTATTAAGTGTCAAGTGTCTTTGC 222  
DB 164 CGTAATTAATGCTATAGG--GAGTGTCTGTATTAAGTGTCAAGTGTCTTTGC 222  
QY 238 GACATTGTGCGACACCGTGGCGCTATTAGGATATATAGCGAGTGAAGCGAGGAT 297  
DB 238 GACATTGTGCGACACCGTGGCGCTATTAGGATATATAGCGAGTGAAGCGAGGAT 297  
QY 223 GACATTGTGCGACACCGTGGCGCTATTAGGATATATAGCGAGTGAAGCGAGGAT 282  
DB 223 GACATTGTGCGACACCGTGGCGCTATTAGGATATATAGCGAGTGAAGCGAGGAT 282  
QY 298 CTCCATTG--ACCGGAAATTTGAACGAGCAGCGCATGCGGGCTTCTACGAGATCG 356  
DB 298 CTCCATTG--ACCGGAAATTTGAACGAGCAGCGCATGCGGGCTTCTACGAGATCG 356  
QY 283 CTCCATTG--ACCGGAAATTTGAACGAGCAGCGCATGCGGGCTTCTACGAGATCG 342  
DB 283 CTCCATTG--ACCGGAAATTTGAACGAGCAGCGCATGCGGGCTTCTACGAGATCG 342  
QY 357 TGATCAAGGTGCGAGCGAGCTGAGCAGCAGCTGCGGGCATTTCTGCTGTTTGA 416  
DB 357 TGATCAAGGTGCGAGCGAGCTGAGCAGCAGCTGCGGGCATTTCTGCTGTTTGA 416

DB 343 TGATTAAGTCCCGAGCGACCTTGAACGGGCTCTGCGCGGCAATTTCTGACAGCTTTGTGA 402  
QY 417 GCTGGTGGCCGAGAGGAATGGAGCTGCCCCCGGATTTGACATGATCTGAATCTGA 476  
DB 403 ACTGGTGGCCGAGAGGAATGGAGTGGCCCGGATTTGACATGATCTGAATCTGA 462  
QY 477 TTGAGCAGCACCCTGACCGTGGCGGAGAGCTGACGCGGCTTCTGCTGCAATGCG 536  
DB 463 TTGAGCAGCACCCTGACCGTGGCGGAGAGCTGACGCGGCTTCTGCTGCAATGCG 522  
QY 537 GCCCGTGAAGAGCCCGGAGGCGCTCTTCTTGTTCAGTTGAGAGGCGAGTCT 596  
DB 523 GCCCGTGAAGAGCCCGGAGGCGCTCTTCTTGTTCAGTTGAGAGGCGAGTCT 582  
QY 597 ACTTCCACTCCATATTTCTGTGTGACACCAAGGGGTCAATTCATGCTGCGCGCT 656  
DB 583 ACTTCCACTCCATATTTCTGTGTGACACCAAGGGGTCAATTCATGCTGCGCGCT 642  
QY 657 TCCTGAGTCAAGTTAGGAGACAGCTGTGACAGACCATCTACCGGGGATGAGCGGACCC 716  
DB 643 TCCTGAGTCAAGTTAGGAGACAGCTGTGACAGACCATCTACCGGGGATGAGCGGACCC 702  
QY 717 TGCCCACTGCTTGGCGGTGACCAAGCGCTAATGCGCGGAGGGGGAACAAGGTGG 776  
DB 703 TGCCCACTGCTTGGCGGTGACCAAGCGCTAATGCGCGGAGGGGGAACAAGGTGG 762  
QY 777 TGAGAGTGTCTACATCCCACTAATCTCTGCGGAGCTGAGCTGAGTGG 836  
DB 763 TGAGAGTGTCTACATCCCACTAATCTCTGCGGAGCTGAGCTGAGTGG 822  
QY 837 CGTGACTAATGAGAGGATATTAAGCGCTGTTGAACCTGCGCGGAGCGCAACGCG 896  
DB 823 CGTGACTAATGAGAGGATATTAAGCGCTGTTGAACCTGCGCGGAGCGCAACGCG 882  
QY 897 TCGTGGCGGACCACTGACCCCACTGACCGGAGCGGAGGAGGAGGAGATCTGA 956  
DB 883 TCGTGGCGGACCACTGACCCCACTGACCGGAGCGGAGGAGGAGGAGATCTGA 942  
QY 957 ACCCAATTTGACGCGCGCTGTCTAATCGGCTCAAAAACCTCGCGCGCTACATGAGCTGG 1016  
DB 943 ACCCAATTTGATGCGCGCGCTGTCTAATCGGCTCAAAAACCTCGCGCGCTACATGAGCTGG 1002  
QY 1017 TCGGCTGCTGTGAGCGCGGCTCACTCCGAGAGCAGTGTGATCCAGAGGAGCAGG 1076  
DB 1003 TCGGCTGCTGTGAGCGCGGCTCACTCCGAGAGCAGTGTGATCCAGAGGAGCAGG 1062  
QY 1077 CCTGTGATCTCTTCAACGCGCTTCAACTGCGGCTCCAGATCAAGCGCGCTGTGG 1136  
DB 1063 CCTGTGATCTCTTCAACGCGCTTCAACTGCGGCTCCAGATCAAGCGCGCTGTGG 1122  
QY 1137 ACAATGCGGCAAGATCATGCGGCTGACCAATCCGCGCGGCTACCTGTAGGCGCCG 1196  
DB 1123 ACAATGCGGCAAGATCATGCGGCTGACCAATCCGCGCGGCTACCTGTAGGCGCCG 1182  
QY 1197 CTCCGCGCGGACATTAACCAACCGCATCTACCGCATCTGAGAGTGAACGCGTACG 1256  
DB 1183 CTCCGCGCGGACATTAACCAACCGCATCTACCGCATCTGAGAGTGAACGCGTACG 1242  
QY 1257 AACCTGCTACGCGCGCTCGCTCTTCTGCGGCTGCGGCGGAGGAGGAGCGCA 1316  
DB 1243 AACCTGCTACGCGCGCTCGCTCTTCTGCGGCTGCGGCGGAGGAGGAGCGCA 1302  
QY 1317 ACACACTGTGCTGTTGGCGCGGCAACCGGCAAGCAACATCGCGGAGGAGGAGGATCG 1376  
DB 1303 ACACACTGTGCTGTTGGCGCGGCAACCGGCAAGCAACATCGCGGAGGAGGAGGATCG 1362  
QY 1377 CCCAGCGGCTGCTTCTACGCGCTGCTCACTGAGCAATGAGAACTTCCCTTCAATG 1436  
DB 1363 CCCAGCGGCTGCTTCTACGCGGCTGCTCACTGAGCAATGAGAACTTCCCTTCAATG 1422  
QY 1437 ATTGCGTGAACAAGATGCTGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGTCTGG 1496  
DB 1423 ATTGCGTGAACAAGATGCTGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGTCTGG 1482

QY	1497	AGTCCGCCAAGGCCATTCTCGGCGGCAGCAAGGTGCGCTGGACCAAAAGTGCAAGTCGT	1556
Db	1483	AGTCGGC <del>C</del> AAAGCCATTCCTCGGAGGAAGCAAGGTGCGGTGGACCAAAATGCAAGTCCT	1542
QY	1557	CCGCCAGATGACCCCAACCCCGTGATCGTCACTCCAAACCAACATGTGCGCGTGA	1616
Db	1543	CGGCCAGATAGACCCGACTCCCGTGATCGTCACTCCAAACCAACATGTGCGCGTGA	1602
QY	1617	TTGACGGGAACAGCACCACTTCGAGCACAGCAGCCGTTGCAGGACCCGATGTTCAAT	1676
Db	1603	TTGACGGGAACCTCAACGACTTCGAAACACAGCAGCCGTTGCAAGACCGGATGTTCAAT	1662
QY	1677	TTGAACTCACCCCGCGTCTGGAGCATGACTTTGGCAAGGTGACAAAGCAGGAAGTCAAG	1736
Db	1663	TTGAACTCACCCCGCGTCTGGATCATGACTTTGGAGAGTTCACCAAGCAGGAAGTCAAG	1722
QY	1737	AGTTCCTCCGCTGGGCGCAGATCACTGACCGAGTGGCGCATGAGTTCTACGTCAAGAA	1796
Db	1723	ACTTTTCCGCTGGGCAAGGATCACGTGTTGAGGTGAGCATGAATTTACGTCAAAA	1782
QY	1797	AGGGTGGAGCCAAACAAGACCCGCCGATGACGCGGATAAAGCGAGCCCAAGCGGG	1856
Db	1783	AGGGTGGAGCCAAAGAAGAAGACCCGCCCAAGTACGCGAGATATAAGTAGCCCAACGGG	1842
QY	1857	CCTGCCCTCAGTCCGCGATCCATCGACGTACAGCCGGAAGAAGCTCCGCTGACTTTG	1916
Db	1843	TGCGCGAGTCAGTTCGCGACCCATCGACGTACAGCCGGA--AGCTTCGATCAACTACG	1899
QY	1917	CCGACAGGTACCAAAACAATGTTCTCGTCAACGCGGAGTCTTCAGATGCTGTTCCCT	1976
Db	1900	CAGACAGGTACCAAAACAATGTTCTCGTCAACGTGGGATGAATCTGATGCTGTTCCCT	1959
QY	1977	GCAAGACATGCGAGAGATGAATCAGAATTTCAACATTGCTTACGCACGCGGACGAGAG	2036
Db	1960	GCAGACAATGCGAGAGATGAATCAGAATTTCAACATTGCTTCACTCACGCGACAGAAG	2019
QY	2037	ACTGTTCAAGTGTCTCCCGCGGTGTCAGAATCTCAACCGGT--CGTCAGAAAGAGA	2093
Db	2020	ACTGTTAAGTGTCTTCC--CGTGTCAGAATCTCAACCCGTTTCTGTCTGTCAAAAGG	2076
QY	2094	CGTATCGAAACTCTGTGCCATTATCATCTGCTGGGCGGGCTCCCGAGATTGCTTGTCT	2153
Db	2077	CGTATCAGAACTGTGTCAATTCATTCATATCATGTGGAAAGGTGCCAGA--CGCTTGCA	2133
QY	2154	CGGCTGCGATCTGTGTCACTGTGAACCTGTGACTGTGTTCTGAGCAATTAATGACTT	2213
Db	2134	CTGCTGCGATCTGTGTCAATGTGATTGTGATGACTGCACTTTGAACAATAATGATTT	2193
QY	2214	AAACCAGGTATGCTGCCGATGTTATCTTCCAGATTGGCTCGAGACAACCTCTCGAG	2273
Db	2194	AAATCAGGTATGCTGCCGATGTTATCTTCCAGATTGGCTCGAGACAACCTCTCTGAA	2253
QY	2274	GGCATTCCGCGAGTGTGGACTTGAACCTTGAGCCGCCAAGCCCAAGCCCAACCAAGCAA	2333
Db	2254	GGAATTAAGACAGTGTGTGAAGCTCAAACTGGCCCAACCAACCAAGCCCGCAGAGCGG	2313
QY	2334	AAGCAGGACGACCGCGCGGTCTGTGCTTCTGTGCTTCAAGTACCTCGGACCTTCAAC	2393
Db	2314	CATAAGGACGACAGCGGGGTCTGTGCTTCTGTGCTTCAAGTACCTCGGACCTTCAAC	2373
QY	2394	GGAATCGACAAGGGGAGCCCGTCAACGCGCGGACGACGCGCCCTCGAGCAGCAAG	2453
Db	2374	GGAATCGACAAGGGGAGCCCGTCAACGAGCAGACGCGCGCCCTCGAGCAGCAAAA	2433
QY	2454	GCCTACGACGACGCTCAAAGCGGCTGACAATCCGTACTGCGGTATAACACGCGGAC	2513
Db	2434	GCCTACGACCGGCAAGCTCGACAGCGGAGACAACCCGTACTCAAGTACAACACGCGGAC	2493
QY	2514	GCCGAGTTTCAAGAGCGCTTGCAAGAAGTACGTCTTTTGGGGGCAACCTCGGGCGACA	2573
Db	2494	GCCGAGTTTCAAGAGCGCTTTAAAGAAGTACGTCTTTTGGGGGCAACCTCGGACGACCA	2553

QY	2574	GTCTTCCAGGCCAAGAAGCGGGTTC	TCGAACCTCTCGGTCTGTGTGAGGAAGCGCTAAG	2633
Db	2554	GTCTTCCAGGCGAAAAAGAGGGTTC	TGAACCTCTGGGCTGTGTGAGGAACCTGTTAAG	2613
QY	2634	ACGGCTCCTGGAAAGAAACGTCCGGTA	GAGCAGTCCGCACAAGAGCCAGACTCCTCTCG	2693
Db	2614	ACGGCTCCGGAAAAAAGAGCGCGTAGA	GCATCTCTCTGTGAGCCAGACTCCTCTCG	2673
QY	2694	GGCATCGCAAGACAGGCCAGCAGCCCG	CTAAAGAGACTCAATTTTGTTCAGACTGGC	2753
Db	2674	GGAACCGGAAAGCGGGCCAGCAGCCTG	CAAGAAAAAGATTGAAATTTGTTCAGACTGGA	2733
QY	2754	GACTCAGAGTCAGTCCCCGATCCACA	CCCTCTCGGAGAACTCCAGCAACCCCGCTGCT	2813
Db	2734	GACGCAAGCTCAGTACTGAACTGACCC	CAAGCCTCTTCGGAACAAGCACCAGCAGCCCTCTGTGT	2793
QY	2814	GTGGGACCTACTACAAATGGCTTCAGG	CGGTGGCGCAACCAATGGCAGACATAACGAAGGC	2873
Db	2794	CTGGAACTAATACGATGGCTACAGGC	AGTGGCGCACCAATGGCAGACATAACGAAGGC	2853
QY	2874	GCCGACGGAGTGGGTAATGCTCAGAA	ATTGGCATTGGCATTCACATGGCTGGGCGAC	2933
Db	2854	GCCGACGGAGTGGGTAATTCCTCCGAA	ATTGGCATTGGCATTCACATGGATGGGCGAC	2913
QY	2934	AGAGTCATCACCAACGACACCCGACCT	GGCCTTGCCCACTACAATTAACCACTCTAC	2993
Db	2914	AGAGTCATCACCAACGACACCCGAACT	GGGCGCTGCCCACTACAACAACCACTCTAC	2973
QY	2994	AAGCAATCTCCAGTGCTTCAACGGGG	CGACGAACAACCACTACTTCGGCTACAGC	3053
Db	2974	AAACAATTTCCAGCCAATCA--GGA	GCTTCGAACGACAATCACTATTGGCTACAGC	3030
QY	3054	ACCCCTGGGGTATTTTGTATTTCAA	CAAGATTCCACTGCCACTTTTCAACCACTGACTGG	3113
Db	3031	ACCCCTGGGGGTATTTTGACTTCAAC	AGATTCCACTGCCACTTTTCAACCACTGACTGG	3090
QY	3114	CAGGACTCATCAACAACAATTGGGG	ATTCCGGCCCCAAGAGACTCAACTTCAAACTCTTC	3173
Db	3091	CAAGACTCATCAACAACAATGGGATT	CCGACCCCAAGAGACTCAACTTCAAACTCTTT	3150
QY	3174	AACATCCAAGTCAAGGAGTCAACGAA	TGATGGCTCAACAACCATCGTAATAACTT	3233
Db	3151	AACATTCAGTCAAGAGAGTCAACGAA	TGACGGTACGAGCATGGTCCAATAAATCTT	3210
QY	3234	ACCAGCAGGTTCAAGTCTTCTCGA	CTCGGAGTACCAAGCTTCGTACGTCCTCGGCTCT	3293
Db	3211	ACCAGCAGGTTCAAGTGTCTTACTGA	CTCGGAGTACCAAGCTTCGTACGTCCTCGGCTCG	3270
QY	3294	GCGCACCAGGGGCTGCTCCCTCCGTT	CCCGCGGACGTGTTCAATGCCCAATAACGGC	3353
Db	3271	GCGCATCAAGGATGCTCTCCCGCTT	CCGACGACAGCTCTTCAATGGTCCACAGTAATGGA	3330
QY	3354	TACCTGACGCTCAACAATGGCAGAC	CAAGCCGTGGGACGTTCACTCTTTTACTGTGCGAA	3413
Db	3331	TACCTCACCTGAAACAACGGGAGT	CAGGCACTAGGACGCTCTTCAATTTTACTGTGCTGAG	3390
QY	3414	TATTTCCCTTCTCAGATGCTGAGAA	CGGGCAAACTTACCTTACGCTTACACCTTTGAG	3473
Db	3391	TACTTCCCTTCTCAGATGCTGCGTAC	CGGAAACAATTACCTTACGCTTACACTTTTGAG	3450
QY	3474	GAAGTGCTTTTCCAAGAGCTACAGC	GCAGCAGACCAAGCCTGACCGGCTGATGAATCCT	3533
Db	3451	GACGTTCTTTCCAAGAGCTACGCTCA	CAGCAGAGTCTGACCGCTCATGAATCCT	3510
QY	3534	CTCATCGACCAATACTGTATTACTGA	ACAGAACTCAAAATCAGTCCGGAAGTGCCCAA	3593
Db	3511	CTCATCGACCAAGTACTGTATTACTTG	AGCAGAAACAACACTCCAAGTGGAAACCAACACG	3570
QY	3594	AACAAGGACTTGCTGTTTAAGCCGT	GGGTCTCCAGCTGGCATGTCTGTTCAAGCCCAAAAAC	3653
Db	3571	CAGTCAAGGCTTCAGTTTCTCAGG	CCGGAGCGAGTGACATTTCCGGGACCACTTAAGAAAC	3630
QY	3654	TGCTACTGTGACCTGTATTATCGG	CAGCAGCGCGTTTCTAAAACAATAACAGACAACAAC	3713



Db	3631	TGGCTTCCTGGACCCCTGTTACCCGCCAGCAGCAGTATCAAGACATCTGCGATAACAC	3690
QY	3714	AACAGCAATTTTACCTGGACTGGTGTCTTCAAATATTAACCTTAATGGGCGTGAATCCATC	3773
Db	3691	AACAGTGAATACTCGTGGACTGGAGCTACCAAGTACCACCTTAATGGCAGAGACTCTGTG	3750
QY	3774	ATCAACCCCTGGCACTGCTATGGCTTCAACAAGACGACGAGACAAGTTCTTCCCATG	3833
Db	3751	GTEAATCCGGGGCCCGCATGGCAAGCCACAAGACGATGAAGAAAGTTTTTCTTCAG	3810
QY	3834	AGCGGTGTCATGATTTTGGAAAAGAGAGCGCCGAGCTTCAACACTGCATTGGACAAT	3893
Db	3811	AGCGGGGTCTCATCTTTGGGAAGCAAGGCTCAGAGAAAACAATGTGAACATTGAAGAAG	3870
QY	3894	GTCATGATTACAGACGAAGAGAAATTAAGCCACTAACCTGTGGCCACCGAAAGATT	3953
Db	3871	GTCATGATTACAGACGAAGAGAAATCGGAACAACCAATCCCGTGGCTACGAGCAGTAT	3930
QY	3954	GGGACCGTGGCAGTCAATTTCCAGACAGACAGACAGACCCTTGCACCCGAGATGTGCAT	4013
Db	3931	GGTCTGTATCTACCAACCTCCAGAGAGGCCAACAGACAAGCACTACCGAGATGTCAAC	3990
QY	4014	GCTATGGAGCATTACCTGGCATGGTGTGGCAAGATAGACAGTGTACCTGCAGGGTCC	4073
Db	3991	ACACAAGCGCTTCTTCAGGCAATGTCTGGCAGACAGAGATGTGTACCTCAGGGGCC	4050
QY	4074	ATTGGGCCAAAATTCCTCAACAGATGGACACTTTCACCGTCTCCTTATGGCGGC	4133
Db	4051	ATCTGGCAAGATTCACACACGAGACGACATTTTCAACCCCTCCTCCCTCATGGGTGA	4110
QY	4134	TTTGAATCAAGAACCCGCTCTCAGATCCTCATCAAAAACAGCCTGTTCTCTGCGAAT	4193
Db	4111	TTGGAATTAACACCCCTCTCCACAGATTTCTATCAAGACACCCCGTACTGCGAAT	4170
QY	4194	CCTCCGGCGAGTTTTCAGCTACAAAGTTTGTCTTCAATCAACCAATACTCCACAGA	4253
Db	4171	CCTTCGACCACCTTCAGTGGCGCAAGTTTGTCTTCTTCAATCAACAGTACTCCACGGGA	4230
QY	4254	CA-AGTAGTGTGGAATTTGAATGGGAGCTGCAGAAAGAAAACGCAAGCGCTGGAATCC	4312
Db	4231	CACGCTCAGCGTGGAGATCGATGGGAGCTGCAGAAAGAAAACGCAAGCGCTGGAATCC	4290
QY	4313	CGAAGTGCAGTACACATCCAAATTATGCAAAATCTGCCAAGTGTGATTTTACTGTGACAA	4372
Db	4291	CGAAATTCAGTACACTTCCAACTACAACAAGTGTGTAATGTGGAATTAACCGTGAATAC	4350
QY	4373	CAATGACTTATATCTGAGCCTCGCCCATTTGGCAACCCGTTACTTACCCTGCCCTGTA	4432
Db	4351	TAATGGCGGTATTCAGAGCCTCGCCCATTTGGCACACAGTACTGACTCGTAACTGTGA	4410
QY	4433	ATTACGTGTTAATCAATAAACCGGTGATTTGCTTCAAGTTGAACCTTGGTCTCCTGTCT	4492
Db	4411	ATTGCTTGTAAATCAATAAACCGTTAATTGTTGTCAGTTGAACCTTGGTCTCCTGTCT	4470
QY	4493	TCTTATCTTATCGGTTACCATGGTTATAGCTTACACATTAATGCTTGGTGTGCGCTTGC	4552
Db	4471	TCTTCTTATCTAGTTTCCATGGCTACGTAGATTAATGATGCGGGTTAATCAATTA	4530
QY	4553	GATAAAGACTTAAGTCAATCGGGTTACCCCTAGTATGAGTGGCCCACTCCCTCTGTC	4612
Db	4531	CTACAAGA-----ACCCTAGTATGAGTTGGCCACTCCCTCTGTC	4574
QY	4613	GCGCTCGCTCGCTCGGTGGGGCTGCGGACCAAAAGTCCGCAAGCGCAGAGCTCTGCTC	4672
Db	4575	GCGCTCGCTCGCTCACTGAGGCCGGGGCGAACCAAAAGTCCGCCAACCGCGGCTTGGCC	4634
QY	4673	TGCCGGCCCCACCGAGCGAGCGAGCGCGCAGAGAGGAGTGGGCAAA	4718
Db	4635	GGCGGGCCTCAGTAGCGAGCGAGCGCGCAGAGAGGAGTGGGCAAA	4680

```

US-08-475-391-1
; Sequence 1, Application US/08475391
; Patent No. 5786211
; GENERAL INFORMATION:
; APPLICANT: Johnson, Philip R.
; TITLE OF INVENTION: Adeno-Associated Virus Materials and
; TITLE OF INVENTION: Methods
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 S. Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,391
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/254,358
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 578621land, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 31975
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4680 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-475-391-1

Query Match          65.3%; Score 3082.4; DB 1; Length 4680;
Best Local Similarity 80.0%; Pred. No. 0;
Matches 3781; Conservative 0; Mismatches 891; Indels 54; Gaps 11;

QY      1 TTGCCACTCCCTCTGCGCGCTCGTCGCTGGTGGGCGCTGCAGCAACAAAGTCCGC 60
       ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      1 TTGGCCACTCCCTCTGTGCGCGCTCGTCGCTCAGTGAGCGCGGCGAACGAAGSTCGCC 60

QY      61 AGACGGCAGAGCTGTGCTGTGCCCCGCCCCACCAGCGAGCGAAGCGCGCAGAGAGGAGTG 120
       ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      61 CGACGCCCGGGCTTTGGCCCGGGCGGCTCAGTAGCGAGCGAGCGCGCAGAGAGGAGTG 120

QY      121 GGCAACTCCATCACTAGGGGTAATCGCGAAGCGCTCCACGCTGCCGCTCAGCGCTGA 180
       ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      121 GCCAACTCCATCACTAGGGGT-----TCCTGAGGGGTGAGTCGTGA 163

QY      181 CGTAAATTACGTCATAGG---GAGTGTCTGTATTAGCTGTACAGTGAGTCTTTGC 237
       ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      164 CGTGAATTACGTCATAGGGTTAGGAGAGTCTGTATTAGAGGTACAGTGAGTG-TTTTGC 222

QY      238 GACATTTTGCAGACACGATGCGCATTTAGAGGTATATATAGCCGAGTGAGCGAGAGAT 297
       ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      223 GACATTTTGCAGACACCATGTGTGTACAGCTGGGTATTAGCCCGAGTGAGCACGAGGAT 282

QY      298 CTCATTTTG-ACCGCGAAATTGAACGAGCAGCAGCCATGCGGGCTTTCAGAGATCG 356
       ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      283 CTCATTTTGAAGCGGAGGTTTGAACGCGCAGCGCCCATGCGGGGTTTACGAGATTG 342

QY      357 TGATCAAGGTCCGAGCGAAGCTTGACGAGCAGCAGCCTGCGGGCATTTCTGACTCGTTGTGA 416

```

Db	343	TGATTAAGGTCCCCAGCGACCTTGACGGGCATCTGCCCGGCATTTCTGACAGCTTTGTGA	402
QY	417	GCTGGGTGCCCCGAGAAGATGGAGCTGCCCCCCGGATTCTGCATGATCTGAATCTGA	476
Db	403	ACTGGGTGCCCCGAGAAGATGGAGTTGCCCGCAAGATTTCTGCATGGAATCTGA	462
QY	477	TTGAGCAGGCACCCCCTGACCGTGGCCGAGAAGCTGCAGCGCGCATTCCTGGTCCAATGGC	536
Db	463	TTGAGCAGGCACCCCCTGACCGTGGCCGAGAAGCTGCAGCGCGCATTCCTGACGGAATGGC	522
QY	537	GCCGCGTGAAGAGCCCCCGAGAGCCCTCTTTTGTTCAGTTGAGAAAGGCGAGTCTT	596
Db	523	GCCGCGTGAAGAGCCCCCGAGAGCCCTTTCTTTGTGCAATTTGAGAAGGAGAGAGCT	582
QY	597	ACTTCCA <sup>1</sup> CTCCATATTTCTGTTGAGAGACCA <sup>2</sup> CGGGGGTCAATTCATGTGCTGGCCGCT	656
Db	583	ACTTCCACATGCACGTGCTCGTGA <sup>1</sup> AA <sup>2</sup> CCCGGGGTGAATTCATGTGTTTGGGACGTT	642
QY	657	TCCTGAGTCAGATTAGGGACAAGCTGGTGCAGAGCCATCTACCGCGGATCGAGCCGACCC	716
Db	643	TCCTGAGTCAGATTGCGA <sup>1</sup> AAAA <sup>2</sup> CTGATT <sup>3</sup> CAGAGAATTTACCGCGGATCGAGCCGACTT	702
QY	717	TGCCCAACTGGTTTCGCGGTGACCAAGACGCTAATGACGCGCGAGGGGGAACAAGGTGG	776
Db	703	TGCCAAACTGGTTTCGCGGTCA <sup>1</sup> AAAGACCA <sup>2</sup> GAATGCGCGCGAGCGGGAACAAGGTGG	762
QY	777	TGACACAGTGTACATCCCCAACTACCTCTGCCCCAAGCTCACGCCGAGCTGCAGTGGG	836
Db	763	TGAGTAGTGTACATCCCAATTA <sup>1</sup> CTTGCTCCCA <sup>2</sup> AAACC <sup>3</sup> CAGCTC <sup>4</sup> CAGTGGG	822
QY	837	CGTGA <sup>1</sup> CTAACATGAGAGATATTAAGCGCCTGTTGAACCTGGCCGACGCAACCGC	896
Db	823	CGTGA <sup>1</sup> CTAATATGAACAGATATTAAGCGCCTGTTGAATCTCA <sup>2</sup> CGAGCGTAACCGT	882
QY	897	TCGTGGCGCAGCACCTGACCCACGTCAGCCAGACCCAGAGAGCAACAAGAGAATCTGA	956
Db	883	TGCTGGCGCAGCATCTGACGCA <sup>1</sup> GTGTCAGACGCAAGAGCAACAAGAGAATCTGA	942
QY	957	ACCCCAATTTGACGCGCCTGTCA <sup>1</sup> TCCGGTCA <sup>2</sup> AAAA <sup>3</sup> ACCTCCGCGCTACATGAGCTGG	1016
Db	943	ATCCCAATTTGATGCGCGGTGATCA <sup>1</sup> GATCA <sup>2</sup> AAAA <sup>3</sup> CTTCA <sup>4</sup> CGCAGTACATGAGCTGG	1002
QY	1017	TCGGGTGGCTGTGGA <sup>1</sup> CCGGGCA <sup>2</sup> TCACCTCCGAGAAGCAGTGAATCCAGAGACCAAG	1076
Db	1003	TCGGGTGGCTGTGGA <sup>1</sup> CAAGGGGA <sup>2</sup> TTA <sup>3</sup> CTCGAGAAGCAGTGAATCCAGAGACCAAG	1062
QY	1077	CCTCGTACATCTCCTTCAACGCGGCTTCCA <sup>1</sup> ACTGCGGTCCAGATCAAGCGGCTTGG	1136
Db	1063	CCTCATACATCTCCTTCAATGCGGCTCCA <sup>1</sup> ACTGCGGTCCCA <sup>2</sup> AATCAAGGCTGCTTGG	1122
QY	1137	ACAATGCCGGAAGATCATGCGGCTGA <sup>1</sup> CCAAATCCGCGCCGCACTA <sup>2</sup> CTGTGAGCCCCG	1196
Db	1123	ACAATGCCGGAAGATTATGAGCTGA <sup>1</sup> AA <sup>2</sup> AA <sup>3</sup> CCGCCCGCACTA <sup>4</sup> CTGTGTGGCCAGC	1182
QY	1197	CTCGCCCCGGACATTAA <sup>1</sup> AA <sup>2</sup> CAACCGCATCTACCGCATCTGAGCTGAACGGCTACG	1256
Db	1183	AGCCCGTGAGGACATTTCAGCAATCGGATTAT <sup>1</sup> AAATTTGA <sup>2</sup> A <sup>3</sup> CTAAACGGGTACG	1242
QY	1257	AACTGCTACGCGGCTCCGCTTTCTCGGCTGGGCCCA <sup>1</sup> AAAGTTCCGGAAGCGCA	1316
Db	1243	ATCCCAATATGCGGCTTCCGCTTTCTGGGATGGCCACGA <sup>1</sup> AAAGTTCCGCAAGAGA	1302
QY	1317	ACACCATCTGGCTGTTTGGGCGCGCA <sup>1</sup> CCACGGGCAAGACCA <sup>2</sup> CATCGCGGAAGCCATCG	1376
Db	1303	ACACCATCTGGCTGTTTGGGCGCTGCA <sup>1</sup> CTACCGGGAAGACCA <sup>2</sup> CATCGCGGAAGCCATAG	1362
QY	1377	CCCAAGCCGTGCCCTTCTACGGCTGCGTCA <sup>1</sup> ACTGGA <sup>2</sup> CCAATGAACTTCCCTTCAATG	1436
Db	1363	CCCACACTGTGCCCTTCTACGGGTGCGTAA <sup>1</sup> CTGA <sup>2</sup> CCAATGAACTTCCCTTCAACG	1422
QY	1437	ATTGCGTGAACAAGATGTGATCTGTGGAGGAGGCAAGATGACGGCCAAGGTCTGTG	1496

D	b		1423	A	C	T	G	T	G	T	G	C	A	A	G	A	T	G	T	G	T	G	T	G	A	G	G	G	A	A	G	A	T	G	A	C	C	G	C	A	A	G	T	G		1482							
O	y		1497	A	G	T	C	C	G	C	A	A	G	S	C	A	T	T	C	T	C	G	G	C	G	C	A	G	C	A	A	G	T	G	C	C	G	T	G	A	C	C	A	A	A	A	G	T	G		1556		
D	b		1483	A	G	T	C	G	G	C	C	A	A	A	G	C	A	T	T	C	T	C	G	G	A	G	A	G	C	A	A	G	T	G	C	C	G	T	G	A	C	C	A	A	A	T	G		1542				
O	y		1557	C	C	G	C	C	C	A	G	A	T	C	G	A	C	C	C	C	C	G	T	G	A	T	C	G	T	C	A	C	C	T	C	C	A	A	C	A	C	A	T	G	T	G	C	C	G	T	G		1616
D	b		1543	C	G	G	C	C	C	A	G	A	T	A	G	A	C	C	C	G	A	C	T	C	C	G	T	G	A	T	C	G	T	C	A	C	C	A	C	A	C	A	T	G	T	G	C	C	G	T	G		1602
O	y		1617	T	T	G	A	C	G	G	A	C	A	G	C	A	C	C	T	T	C	G	A	C	A	C	C	A	G	C	A	G	C	C	G	T	T	G	C	A	G	A	C	C	G	A	T	G	T		1676		
D	b		1603	T	T	G	A	C	G	G	A	C	T	C	A	C	C	T	T	C	G	A	C	A	C	C	A	G	C	A	G	C	C	G	T	T	G	C	A	A	G	A	C	C	G	A	T	G	T		1662		
O	y		1677	T	T	G	A	C	T	C	A	C	C	C	C	G	C	T	T	G	A	G	C	A	T	G	A	C	T	T	T	G	C	A	A	G	T	G	A	C	A	A	G	C	A	A	G	T	G		1736		
D	b		1663	T	T	G	A	C	T	C	A	C	C	C	C	C	G	C	T	T	G	A	T	C	A	T	G	A	C	T	T	G	G	A	A	G	T	C	A	C	C	A	A	G	C	A	A	G		1722			
O	y		1737	A	G	T	T	C	T	C	G	C	T	G	G	C	G	A	G	A	T	C	A	C	G	T	G	A	C	C	G	A	G	T	G	C	C	A	T	G	A	G	T	T	C	T	A	C	G	T		1796	
D	b		1723	A	C	T	T	T	T	C	C	G	T	G	G	C	A	A	G	A	T	C	A	C	G	T	G	T	T	G	A	G	T	G	A	G	C	A	T	G	A	T	T	C	T	A	C	G	T		1782		
O	y		1797	A	G	G	T	G	A	G	C	C	A	C	A	A	A	G	A	C	C	C	C	C	C	G	A	T	G	A	C	G	C	G	A	T	A	A	A	G	C	G	A	C	C	C	A	A	G	C	G		1856
D	b		1783	A	G	G	T	G	A	G	C	C	A	A	A	A	A	G	A	C	C	C	C	C	C	C	A	G	T	G	A	C	G	A	T	A	T	A	G	T	A	G	C	C	C	A	A	C	G		1842		
O	y		1857	C	C	T	G	C	C	C	C	T	C	A	G	T	C	C	G	A	T	C	C	A	G	C	T	C	A	G	C	G	A	A	G	A	G	A	G	A	G	A	G	C	T	C	C	G	T	G		1916	
D	b		1843	T	G	C	G	C	G	A	G	T	G	C	G	A	C	C	A	T	C	G	A	C	G	T	C	A	G	C	G	C	G	A	--	A	G	C	T	T	C	G	A	T	C	A	C	T	A	C	G		1899
O	y		1917	C	C																																																

QY 2574 GTCTCCAGGCCAAGAGCGGGTCTCGAAGCTCTCGGTCTGGTTGAGAAAGCGCTAAG 2633  
|||  
Db 2554 GTCTTCCAGGCGGAAAAAGAGGGTCTTGAAGCTCTGGGCTGTGTGAGAACTGTAAAG 2613  
|||  
QY 2634 ACCGCTCCTGGAAGAAACGTCCGGTAGAGAGCTGCCACAAGAGCCAGACTCCTCTCG 2693  
|||  
Db 2614 ACCGCTCCGGGAAAAAGAGGCGGTAGAGCACTCTCTGTGAGGCCAGACTCCTCTCG 2673  
|||  
QY 2694 GGCATCGGCAAGACAGGCCAGCAGCCCGCTAAAGAGACTCAATTTGGTCAAGCTGGC 2753  
|||  
Db 2674 GGAACCGGAAGCGGGCCAGCAGCCTGCAAGAAAAAGATTGAATTTGGTCAAGCTGGA 2733  
|||  
QY 2754 GACTCAGAGTCAAGTCCCGATCCACAACCTCTCGGAGAACTCCAGCAACCCCGTGTCT 2813  
|||  
Db 2734 GACGCAAGACTCAAGTACCTGACCCCAAGCTCTCGGACAGCCACAGAGCCCTCTGTGT 2793  
|||  
QY 2814 GTGGACCTACTAATGAGCTTCAAGCGGTGGGCAACATGACAGCAATAAGAGGC 2873  
|||  
Db 2794 CTGGGAATTAATACGATGGCTACAGGCAAGTGGGCAACATGACAGCAATAAGAGGC 2853  
|||  
QY 2874 GCCGACGAGTGGGTAATGCTCTCAGGAATTTGGCAATTGGCATTCAGATGGCGGAC 2933  
|||  
Db 2854 GCCGACGAGTGGGTAATTCCTCCGGAATTTGGCAATTGGCATTCAGATGGCGGAC 2913  
|||  
QY 2934 AGAGTCATCACCAACAGACACCCGCACTGGGCTTGGCCCACTCAATTAACCACTCTAC 2993  
|||  
Db 2914 AGAGTCATCACCAACAGACACCCGCACTGGGCTTGGCCCACTCAATTAACCACTCTAC 2973  
|||  
QY 2994 AAGCAATCTCCAGTCTTCAACGGGGCCAGCAAGCAACCACTACTTGGCTACAGC 3053  
|||  
Db 2974 AACAATTTCCAGCAATCA--GAGGCTCGAAGCAATCACTACTTGGCTACAGC 3030  
|||  
QY 3054 ACCCCCTGGGTAATTTGATTTCAACAGATTCACTGCCACTTTTACCAAGTGAAGTGG 3113  
|||  
Db 3031 ACCCCTGGGGTAATTTGACTTCAACAGATTCCAAGTCTTCAACCAAGTGAAGTGG 3090  
|||  
QY 3114 CAGGACTCATCAACAACAATTGGGATTCGGGCCAAGAGACTCAACTCAAACTCTTC 3173  
|||  
Db 3091 CAAAGACTCATCAACAACAAGTGGGATTCGCAACCAAGAGACTCAACTCAAACTCTTC 3150  
|||  
QY 3174 AACATCCAAGTCAAGAGGTCAAGCAGATGATGGCGTCAACAACCATGCTAATAACTT 3233  
|||  
Db 3151 AACATCCAAGTCAAGAGGTCAAGCAGATGATGGCGTCAACAACCATGCTAATAACTT 3210  
|||  
QY 3234 ACCAGCAGGTTCAAGTCTTCTCGGACTCGAGTACAGCTTCCGTACGTCTGGCTCT 3293  
|||  
Db 3211 ACCAGCAGGTTCAAGTCTTCTCGGACTCGAGTACAGCTTCCGTACGTCTGGCTCT 3270  
|||  
QY 3294 GCGCACCAGGCTGCTCCCTCCGTTCCCGGCGAGCTGTTCATGATTCGCAATACGGC 3353  
|||  
Db 3271 GCGCATCAAGAGTGCCTCCGCTCCGTTCCAGCAGAGCTTTCATGATTCGCAATACGG 3330  
|||  
QY 3354 TACCTGACGTTCAACAATGGCAGCCAGCGGTGGAGCTTCACTCTTTACTGCTGGA 3413  
|||  
Db 3331 TACCTGACGTTCAACAATGGCAGCGGTGGAGCTTCACTCTTTACTGCTGGA 3390  
|||  
QY 3414 TATTTCCCTTCTCAGATGCTGAGAAGCGGCAACAATTACCTTCAAGTACCTTTGAG 3473  
|||  
Db 3391 TACTTTCTTCTCAGATGCTGCGTACCGGAAACAATTACCTTCAAGTACCTTTGAG 3450  
|||  
QY 3474 GAAGTGCCTTTCCAGAGAGCTACGCGCAGAGCCAGAGCTTGAACGGGCTGATGATCT 3533  
|||  
Db 3451 GAGGTTCTTTCCAGAGAGCTACGCTCAGAGCCAGAGTCTGAGCCGTCTCATGATCTCT 3510  
|||  
QY 3534 CTATCGACCAATACCTGTATTAAGTGAACAGAACTCAAAATCAGTCCGGAAGTCCCAA 3593  
|||  
Db 3511 CTATCGACCAATACCTGTATTAAGTGAACAGAACTCAAAATCAGTCCGGAAGTCCCAA 3570  
|||  
QY 3594 AACCAAGAGTGTGTTAGCCGTGGGTCTCAGCTGGCATGTCTGTTCAAGCCCAAAAC 3653  
|||  
Db 3571 CAGTCAAGGCTTCAAGTTTCTCAGGCGGAGCGAGTGAATTCGGGACCAAGTCTAAGAAC 3630  
|||

QY 3654 TGGCTACCTGGACCCCTGTATTCGGCAGCAGCGCGTTTCTAAAACAAAACAGACAAAC 3713  
|||  
Db 3631 TGGCTTCTGGACCCCTGTATTCGGCAGCAGCGAGTATCAAGACATCTGCGGATTAACA 3690  
|||  
QY 3714 AACAGCAATTTTACCTGAGTGGTCTTCAAAATATTAAGTGGCGTGAATCCATC 3773  
|||  
Db 3691 AACAGTGAATTAAGTGGAGTGAAGTCAACAGTACCACTCAATGGCAGAGACTCTCTG 3750  
|||  
QY 3774 ATCAACCTGGCACTGTATGGCTTCAACAAGAGCAGCAAGCAAGTCTTCCATG 3833  
|||  
Db 3751 GTGAATCCGGGCGCCGCAATGGCAAGCCACAAGAGCATGAAGAAAGTTTTCCTCAG 3810  
|||  
QY 3834 AGCGGTGTATGATTTTGAAGAAAGAGCGCGGAGCTTCAACAAGTGAATGACAT 3893  
|||  
Db 3811 AGCGGTGTATGATTTTGAAGAAAGAGCGCGGAGCTTCAACAAGTGAATGACAT 3870  
|||  
QY 3894 GTCAATTAACAGAGAGAAATTAAGCCACTAACCTGTGGCCACCGAAAGATT 3953  
|||  
Db 3871 GTCAATTAACAGAGAGAAATTAAGCCACTAACCTGTGGCCACCGAAAGATT 3930  
|||  
QY 3954 GGGACCGTGGCAATTAATTTCCAGAGCAGCAGCAGACCTTGGCAGCGAGATGTGAT 4013  
|||  
Db 3931 GGTTCGTATCTACCAACCTCCAGAGGCAACAGACAGAGCTAACCGAGATGTCAAC 3990  
|||  
QY 4014 GCTATGGAGCATTAACCTGGCATGTGTGGCAAGATAGAGCTGTACCTGAGGTTCC 4073  
|||  
Db 3991 ACACAAGCGTCTTCCAGGCAATGTGTGGCAGAGAGATGTATCTTCAAGGGGCC 4050  
|||  
QY 4074 ATTTGGCCAAATTTCTCTCAACAGATGACACTTTCAACCGTCTCTCTTATGGCGGC 4133  
|||  
Db 4051 ATTTGGCCAAATTTCTCTCAACAGATGACACTTTCAACCGTCTCTCTTATGGCGGC 4110  
|||  
QY 4134 TTTGACTCAAGAACCCGCTCTCAGATCTCATCAAAACAAGCTGTCTCTGCAAT 4193  
|||  
Db 4111 TTTGACTCAAGAACCCGCTCTCAGATCTCATCAAAACAAGCTGTCTCTGCAAT 4170  
|||  
QY 4194 CCTCCGGCGAGTTTCAAGTCAAAAGTTTCTCATTCATCAACCAATACTCCACAGGA 4253  
|||  
Db 4171 CCTCCGGCGAGTTTCAAGTCAAAAGTTTCTCATTCATCAACCAATACTCCACAGGA 4230  
|||  
QY 4254 CA-AGTGTGTGAAATTTGAATGGAGCTGCAGAAAGAAACAAGAGCTGGAATCC 4312  
|||  
Db 4231 CAGGTGTGTGAAATTTGAATGGAGCTGCAGAAAGAAACAAGAGCTGGAATCC 4290  
|||  
QY 4313 CGAAGTGTGAAATTTGAATGGAGCTGCAGAAAGAAACAAGAGCTGGAATCC 4372  
|||  
Db 4291 CGAAGTGTGAAATTTGAATGGAGCTGCAGAAAGAAACAAGAGCTGGAATCC 4350  
|||  
QY 4373 CAATGACTTATTAAGTGTGAGCTGCGCCATTTGGCAGCCGCTTACCCGCTGTA 4432  
|||  
Db 4351 TAATGAGTGTATTAAGTGTGAGCTGCGCCATTTGGCAGCCGCTTACCCGCTGTA 4410  
|||  
QY 4433 ATTAAGTGTATTAAGTGTGAGCTGCGCCATTTGGCAGCCGCTTACCCGCTGTA 4492  
|||  
Db 4411 ATTAAGTGTATTAAGTGTGAGCTGCGCCATTTGGCAGCCGCTTACCCGCTGTA 4470  
|||  
QY 4493 TCTTATTTATTCGTTTACCATGTTATGACTTACACATTAATCTTGGTGGCTTGGC 4552  
|||  
Db 4471 TCTTATTTATTCGTTTACCATGTTATGACTTACACATTAATCTTGGTGGCTTGGC 4530  
|||  
QY 4553 GATTAAGACTTAAGTGTGAGCTGCGCCATTTGGCAGCCGCTTACCCGCTGTA 4612  
|||  
Db 4531 CTACAAGG-----ACCCCTAGTGTGAGTGGCCTCTCTCTG 4574  
|||  
QY 4613 GCGCTGCTGCTCGGTGGGCTGCGGACCAAGGTCCGCAAGCGGAGAGCTGTGCTC 4672  
|||  
Db 4575 GCGCTGCTGCTCGGTGGGCTGCGGACCAAGGTCCGCAAGCGGAGAGCTGTGCTC 4634  
|||  
QY 4673 TGCCGGCCCAACCGAGCGAGCGGAGAGAGGAGTGGGCAA 4718  
|||  
Db 4635 GGGCGGCTTCAAGTGAAGCGAGCGGAGAGAGAGGAGTGGGCAA 4680  
|||



RESULT 4

US-08-709-609-1  
; Sequence 1, Application us/08709609  
; Patent No. 5858775  
; GENERAL INFORMATION:  
; APPLICANT: Johnson, Philip R.  
; TITLE OF INVENTION: Adeno-Associated Virus Materials and  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 6300 Sears Tower, 233 S. Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/709,609  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: No. 5858775and, Greta E.  
; REGISTRATION NUMBER: 35,302  
; REFERENCE/DOCKET NUMBER: 31975  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (312) 474-6300  
; TELEFAX: (312) 474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4680 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-709-609-1

Query Match 65.3%; Score 3082.4; DB 2; Length 4680;  
Best Local Similarity 80.0%; Pred. No. 0;  
Matches 3781; Conservative . 0; Mismatches 891; Indels 54; Gaps 11;

QY 1 TTGCCCCCTCTCTCTGCGCGCTCGCTCGCTCGGTGGGCGCTGGCAACCAAGTCCGC 60  
DB 1 TTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGCGCGGCGCAACCAAGTCCGC 60  
QY 61 AGACGGCAGAGCTCTGCTCTGCGCGCGCGCGCGCGAGCGAGCGCGCAGAGAGGAGTG 120  
DB 61 CGACGCCGCGGCTTGGCGCGCGCGCTCACTGAGCGAGCGCGCAGAGAGGAGTG 120  
QY 121 GGCACTTCATCACTAGGGGTAATCGGAAGCGCCTCCACGCTGCGCGCTCAGCGCTGA 180  
DB 121 GCCAATTCATCACTAGGGGTAATCGGAAGCGCCTCCACGCTGCGCGCTCAGCGCTGA 180  
QY 181 CGTAAATTAAGTATAGG---GAGTGGTCTGTATTAGCTGTCAAGTGTGCTTTTGC 237  
DB 164 CGTGAATTAAGTATAGG---GAGTGGTCTGTATTAGCTGTCAAGTGTGCTTTTGC 222  
QY 238 GACATTTTGGCAGACCAAGTGGCCATTAGGGTATATATGCGCGAGTGAAGCGAGAT 297  
DB 223 GACATTTTGGCAGACCAAGTGGCCATTAGGGTATATATGCGCGAGTGAAGCGAGAT 282  
QY 298 CTCCATTTTG-AACCGGAAATTTGAACGAGCAGACGACCATGCGCGGCTTCTACGAGATCG 356  
DB 283 CTCCATTTTGAAGCGGAGGTTTGAACGCGCAGCGCCATGCGCGGCTTCTACGAGATG 342  
QY 357 TGATCAAGGTGCGAGCGAGCCTGAGCAGCAGCCTGCGGCGCATTTCTGACTGTTTGA 416  
DB 343 TGATTAAGGTGCGAGCGAGCCTGAGCAGCAGCCTGCGGCGCATTTCTGACAGCTTTTGA 402

QY 417 GCTGGGTGCGCGAGAGGAATGGAGAGCTGCCCCCGGATTTCTGACATGATCTGAATCTGA 476  
DB 403 ACTGGGTGCGCGAGAGGAATGGAGAGTTCGCCCGCAGATTTCTGACATGATCTGAATCTGA 462  
QY 477 TTGAGCAGGACACCCCTGACCGTGGCCGAGAGAGTGCAGCGCGACTTCCTGTCTCAATGGC 536  
DB 463 TTGAGCAGGACACCCCTGACCGTGGCCGAGAGAGTGCAGCGCGACTTCCTGTCTCAATGGC 522  
QY 537 GCCCGTGAAGTAAAGCCCGGAGGCGCTCTTCTTTGTTGATGTTGAGAGGCGAGTCT 596  
DB 523 GCCGTGTAAGTAAAGCCCGGAGGCGCTCTTCTTTGTTGATGTTGAGAGGCGAGTCT 582  
QY 597 ACTTCCACCTCATATTTCTGGTGAAGACCAAGGCGGTAATCCATGCTGGCGCGCT 656  
DB 583 ACTTCCACATGACAGTGTCTGTGAAGACCAAGGCGGTAATCCATGCTTTGGAGCTT 642  
QY 657 TCCTGAGTCAAGTATAGGAGCAAGCTGTGTCAGACCATCTACCGCGGATCGAGCGGACCC 716  
DB 643 TCCTGAGTCAAGTATAGGAGCAAGCTGTGTCAGACCATCTACCGCGGATCGAGCGGACTT 702  
QY 717 TGCCCAACTGTTGCGCGTGAACCAAGCGTATATGCGCGGAGAGGCGGCAACAGTGG 776  
DB 703 TGCCCAACTGTTGCGCGTGAACCAAGCGTATATGCGCGGAGAGGCGGCAACAGTGG 762  
QY 777 TGAGCAGTCTACATCCCACTACCTCTGCCCCAAGCTCAGCCGAGCTGAGTGG 836  
DB 763 TGAGTAGTCTACATCCCACTACCTCTGCCCCAAGCTCAGCCGAGCTGAGTGG 822  
QY 837 CGTGAATAACATGAGAGATATTAAGCGCCTGTTGAACTTGGCGGAGCGCAACCGC 896  
DB 823 CGTGAATAATATGAAACAGATTAAGCGCCTGTTGAACTTCAACGAGCGTAACCGT 882  
QY 897 TCGTGGCGCAGCAGCCTGACCCAGCTCAGCCAGACCCAGAGCAGAACAGAGATCTGA 956  
DB 883 TCGTGGCGCAGCAGCCTGACCCAGCTCAGCCAGACCCAGAGCAGAACAGAGATCTGA 942  
QY 957 ACCCAATTTGACGCGCGCTGATCCCGTCAAAAACCTTCGCGGCTACATGAGCTGG 1016  
DB 943 ATCCCAATTTGATGCGCGGCTGATCAGTCAAAAACCTTCAGCCAGGTACATGAGCTGG 1002  
QY 1017 TCGGTGGCTGTGAGCCGGGCTACCTCCGAGAGAGTGTCCAGAGAGCAGG 1076  
DB 1003 TCGGTGGCTGTGAGCAAGGGATTAACCTCGAGAGAGTGTCCAGAGAGCAGG 1062  
QY 1077 CCTGTACATCTCTTCAACGCGCTTCCAACTGCGGCTCCAGATCAAGCGCTGG 1136  
DB 1063 CCTGTACATCTCTTCAATGCGGCTTCCAACTGCGGCTCCAAATCAAGCGCTGG 1122  
QY 1137 ACATGCGCGCAAGATCATGCGGCTGACCAATCCGCGCGCACTACTGTAGCGCCG 1196  
DB 1123 ACATGCGCGCAAGATTAATGAGCTGACTAAACCGCGCGCACTACTGTAGCGCGC 1182  
QY 1197 CTCCGCGCGGACATTAACCAACCGCATCTACCGCATCTGAGCTGAACGGCTACG 1256  
DB 1183 AGCCGCTGAGGACATTTCCAGCAATCGGATTTATAAATTTTGAACATAACGGCTACG 1242  
QY 1257 AACCTGCTACGCGGCTCCGCTTCTCTGCGTGGGCCAGAAAAGTTGGGAGCGCA 1316  
DB 1243 ATCCCAATATAGCGGCTTCCGCTTCTGAGTGGGCCAGAAAAGTTGGGAGCGCA 1302  
QY 1317 ACACATCTGCTGTTTGGCGCGCACACCGGCAAGCAACATCGCGGAGCGCATCG 1376  
DB 1303 ACACATCTGCTGTTTGGCGCTGCAACTACCGGCAAGCAACATCGCGGAGCGCATAG 1362  
QY 1377 CCCAGCGGTGCGCTTCTACGCGCTGCTCACTGAGCAATGAGAACTTCCCTCAATG 1436  
DB 1363 CCCAGCTGTGCGCTTCTACGCGGTGCTAACTGAGCAATGAGAACTTCCCTCAACG 1422  
QY 1437 ATGCGTGCAGACAGATGTGATCTGCTGGAGAGGCGCAAGATGACGCGCAAGTCTGG 1496  
DB 1423 ACTGTGTGCAACAAGATGTGATCTGCTGGAGAGGCGCAAGATGACCGCCAAAGTCTGG 1482



Db 3631 TGGCTTCCTGGACCCCTGTACCGCCAGCAGGAGTATCAAGACATCTGCGGATAACAC 3690  
 QY 3714 AACAGCAATTTTACCTGGAAGTGGCTTCAAAATATAACCTCAATGGCGTGAATCCATC 3773  
 Db 3691 AACAGTGAATCTGCTGAGTGAAGTACCAAGTACCACTCAATGGAGAGACTCTCTG 3750  
 QY 3774 ATCAACCTGGCACTGCTATGAGCTTACAGCAAGAGAGAGAGAGAGAGAGAGAGAG 3833  
 Db 3751 GTGAATCCGGGGGGGGCCCATGCGAAGCCACAGACGATGAGAGAAAGTTTTCCTCAG 3810  
 QY 3834 AGCGGTCTCATGATTTTGGAAAAAGAGAGCGCGGAGCTTCAAACTGCTGATGGACAAT 3893  
 Db 3811 AGCGGGGTTCTATCTTTGGGAGCAAGGCTCAGAGAAACAAATGTGAACATTGAAAG 3870  
 QY 3894 GTCATGATTACAGACGAGAGAAATTAAAGCACTAACCTGTGGCCACCGAAAGATT 3953  
 Db 3871 GTCATGATTACAGACGAGAGAAATTAAAGCACTAACCTGTGGCCACCGAAAGATT 3930  
 QY 3954 GGGACCGTGGCAGTCAATTTCCAGAGCAGACAGACCCCTGCGACCGGAGATGTCAT 4013  
 Db 3931 GGTCTGTATCTAACCACTCCAGAGAGGCAACAGCAAGCAGTACCGCAGATGTCAAC 3990  
 QY 4014 GCTATGGAGCATTAACCTGGCATGTGTGGCAAGATAGAGAGTACTGACAGGTTCC 4073  
 Db 3991 ACACAGAGCGCTCTCCAGGATGTGTGGCAAGACAGAGATGTACTTACAGGGGCC 4050  
 QY 4074 ATTTGGCCAAATTCCTCAGACAGATGAGACATTTCAACCCGTCTCTTATGGCGGC 4133  
 Db 4051 ATCTGGCAAGATTTCCACACAGCAGACATTTTCAACCCCTCTCTCTCATGGGTGA 4110  
 QY 4134 TTTGACTCAAGAACCCGCTCTCAGATCTCATCAAAACACAGCCTGTCTCTGCAAT 4193  
 Db 4111 TTGGACTTAAACACCTCTCTCAGATCTCATCAAGAACACCCGCTTACCTGCAAT 4170  
 QY 4194 CCTCCGCGAGTTTTCAGCTACAAAGTTTGTCTCATTCATCACCAATTAATCCACAGA 4253  
 Db 4171 CCTTGCACCACTTCACTGCGCAAGTTTGTCTCTCATCACAGTACTCCACGGA 4230  
 QY 4254 CA-AGTGTGGAATTTGATGAGAGTGCAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 4312  
 Db 4231 CACGTCAGCGTGGAGATCGAGTGGAGTGCAGAAAGAAAGAAAGAAAGCTGGAATCC 4290  
 QY 4313 CGAAGTGCAGTACATCCCAATTAATGCAAAATCTGCCAAGCTGATTTTACTGTGACAA 4372  
 Db 4291 CGAATTCAGTACATTCCTCACTCAACAGTCTGTTAATCTGACTTACCGTGATAC 4350  
 QY 4373 CAATGACTTATTAAGAGCTGCGCCCATTTGGCAACCCGTTACCTTACCCCTGTA 4432  
 Db 4351 TAATGGCGTGTATTCAGAGCTGCGCCCATTTGGCAACAGATACCTGTAATCTGTA 4410  
 QY 4433 ATTACGTGTTAATCAATAAACCGGTTGATTCGTTTCACTTGAATTTGCTCTCTGCT 4492  
 Db 4411 ATTGCTGTGTTAATCAATAAACCGTTAATTCGTTGAGTTGAATTTGCTCTGCTAT 4470  
 QY 4493 TCTTATCTTATCGGTTTACCATGTTATAGCTTACATTAACCTGCTTGGCTTGC 4552  
 Db 4471 TTCTTTCTTATCTAGTTTCCATGCTAGTATGATAATTAAGATGGCGGTTAATCATTA 4530  
 QY 4553 GATAAAGACTTACGTCATCGGGTTACCCCTAGTGTGAGTTGCCACTCCCTCTCTGC 4612  
 Db 4531 CTACAAGG-----ACCCCTAGTGTGAGTTGCCACTCCCTCTCTGC 4574  
 QY 4613 GCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4672  
 Db 4575 GCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4634  
 QY 4673 TGCCGCGCCACGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCG 4718  
 Db 4635 GGGCGGCTCAGTGCAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCG 4680

RESULT 5  
 PCT-US95-07178-1

Sequence 1, Application PC/TUS9507178  
 GENERAL INFORMATION:  
 APPLICANT: Johnson, Philip R.  
 TITLE OF INVENTION: Adeno-Associated Virus Materials and  
 TITLE OF INVENTION: Methods  
 NUMBER OF SEQUENCES: 3  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
 STREET: 6300 Sears Tower, 233 S. Wacker Drive  
 CITY: Chicago  
 STATE: Illinois  
 COUNTRY: USA  
 ZIP: 60606  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent in Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: PCT/US95/07178  
 FILING DATE:  
 CLASSIFICATION:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Noland, Greta E.  
 REGISTRATION NUMBER: 35,302  
 REFERENCE/DOCKET NUMBER: 31975  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (312) 474-6300  
 TELEFAX: (312) 474-0448  
 TELEX: 25-3856  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 4680 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 PCT-US95-07178-1  
 Query Match 65.3%; Score 3082.4; DB 5; Length 4680;  
 Best Local Similarity 80.0%; Pred. No. 0;  
 Matches 3781; Conservative 0; Mismatches 891; Indels 54; Gaps 11;  
 QY 1 TTGCCACTCCCTCTCTGCGCGCTCGCTCGTGGTGGGCGCTGCGACCAAGTCCGC 60  
 Db 1 TTGGCCACTCCCTCTCTGCGCGCTCGCTCGTCACTGAGCGCGGCGGCGACCAAGTCCGC 60  
 QY 61 AGACGGCAGAGCTCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120  
 Db 61 CGACGCGCGCGCTTTGCG 120  
 QY 121 GGCACCTCATCTAGAGGTTATCGCGAAGCGCTCCACGCTCCGCGTCAAGCGCTGA 180  
 Db 121 GCCAATCTCATCTAGAGGTTATCGCGAAGCGCTCCACGCTCCGCGTCAAGCGCTGA 163  
 QY 181 CGTAAATTACGTCATAGG--GAGTGTCTGTATGCTGTCACGTCAGTGTCTTTGC 237  
 Db 164 CGTGAATTACGTCATAGGTTAGGAGGTCCTGTATTAAGGTCACGTAGTG-TTTTGC 222  
 QY 238 GACATTTTGCAGACACGTCGTCATTAGGATATATAGCCGAGTGAGCGAGCAGAT 297  
 Db 223 GACATTTTGCAGACACGTCGTCATTAGGATATTAAGCCGAGTGAGCAGCAGGAT 282  
 QY 298 CTCGATTTTG-ACCGGAATTGAACGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 356  
 Db 283 CTCGATTTTGAAGCGGAGGTTGAACGCGCAGCAGCAGCAGCAGCAGCAGCAGCAG 342  
 QY 357 TGATCAAGGTGCGGAGCGAGCTGGAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 416  
 Db 343 TGATTAAGGTCCCGAGCGACCTTGAAGCGGCGATCTGCGCGGCGATTTCTGACAGCTTTG 402  
 QY 417 GCTGGGTGCGCGAGAGGATGGAGCTGCGCGCGGATTTGACATGATGATCTGAATCTGA 476



Db 403 ACTGGTGCCGAGAAATGGAGTTGCCGCCAGATTCTGACATGATCTGAATCTGA 462  
QY 477 TTGAGCAGCACCCTGACCGTGGCCGAGAACTGTCAGCGGACTTCTGTGTCATGGC 536  
Db 463 TTGAGCAGCACCCTGACCGTGGCCGAGAACTGTCAGCGGACTTCTGTGTCATGGC 522  
QY 537 GCCCGGTGAGTAAGGCCCGGAGGCCCTCTTCTTGTTCAGTTGAGAAAGGCGAGTCT 596  
Db 523 GCCGTGTGAGTAAGGCCCGGAGGCCCTTCTTGTGCAATTGAGAAAGGAGAGAGCT 582  
QY 597 ACTTCCACCTCCATATTTCTGTGAGAGACCAACGGGGTCAAAATCCATGTGTGGCCGCT 656  
Db 583 ACTTCCACATGACGTCGTCTGTGAAACCAACGGGGTGAATCCATGTGTGGGACGTT 642  
QY 657 TCCTGAGTCAGATTAGGACAAAGCTGTGTCAGACCATCTACGCGGATCGAGCCGACCC 716  
Db 643 TCCTGAGTCAGATTGCGAAAACTGATTGAGAAATTACGCGGATCGAGCCGACTT 702  
QY 717 TGCCCACTGTTCCGCGTGAACAAGACGCTAATGCGCGGAGGGGGAACAAGTGG 776  
Db 703 TGCCAACTGTTCCGCGTCACAAGAACAAGATGCGCGGAGGCGGGAACAAGTGG 762  
QY 777 TGAGCAGTCTACATCCCACTACCTCTGCGCCAGACTCAGCCCGAGCTCAGTGGG 836  
Db 763 TGATGAGTCTACATCCCACTACTTCTGCTCCCAAAACCCAGCTGAGCTCCAAGTGG 822  
QY 837 CGTGACTAATGAGAGATATATAAGCGCTGTTGAACCTGCGGAGCGCAACGGC 896  
Db 823 CGTGACTAATGAGACAGATATTAAGCGCTGTTGAATCTCAGGAGCGTAACGGT 882  
QY 897 TCGTGCGCAGCACCCTGACCCACGTCAGCCAGACCCAGAGAGACAAGAAATCTGA 956  
Db 883 TCGTGCGCAGCATCTGACGCAAGCTGTGCAAGCAGAGAGACAAGAAATCAGA 942  
QY 957 ACCCAATCTGACGCGCTGTCTATCCGGTCAAAAACCTCCGCGCTACATGAGCTGG 1016  
Db 943 ATCCCAATCTGATGCGCGGTGATCAGATCAAAAATTCAAGCAGTACATGAGCTGG 1002  
QY 1017 TCGGCTGCTGTGACCGGGGATCACTCCGAGAGAGAGATCCAGAGAGACAGG 1076  
Db 1003 TCGGCTGCTGTGACAGGGGATTAACCTCGAGAGAGAGATCCAGAGAGACAGG 1062  
QY 1077 CCTCGTACATCTCTTCAACGCGCTTCCAACTCGCGGTCCAGATCAAGCGCTTGG 1136  
Db 1063 CCTCATACATCTCTTCAATGCGGCTCCAACTCGCGGTCCCAATCAAGCGCTTGG 1122  
QY 1137 ACAATGCGGCAAGATCAATGCGCTGACCAAAATCCGCGCCGACTACCTGTGAGCCCG 1196  
Db 1123 ACAATGCGGAAAGATTAAGCCTGACTAAACCGCCCGACTACCTGTGAGCCAGC 1182  
QY 1197 CTCCGCGCGGACATTAACCAACCGCATCTACCGCATCTGAGCTGAACGGCTAG 1256  
Db 1183 AGCCCGTGAGGACATTTCCAGCAATCGGATTTATTAATTTTGAACATAACGGGTACG 1242  
QY 1257 AACCTGCTACGCGGCTCGCTCTTCTCGGCTGGGCCAGAAAAAGTTGGGAAGCGCA 1316  
Db 1243 ATCCCAATATGCGGCTTCGCTCTTCTGGGATGGCCAGAAAAAGTTGGCAAGAGA 1302  
QY 1317 ACACCATCTGCTGTTGGGCGGCGCACACCGGGAAGACCAACATCGCGGAGCCATCG 1376  
Db 1303 ACACCATCTGCTGTTGGGCTGCAACTACCGGGAAGACCAACATCGCGGAGCCATAG 1362  
QY 1377 CCCAGCGGCTGCTTCTTCAACGCGTCACTGACCAATGAGAACTTCCCTCAATG 1436  
Db 1363 CCCACACTGTGCCCTTCAACGGGTGCGTAACCTGACCAATGAGAACTTCCCTCAACG 1422  
QY 1437 ATTGCGTCGACAGATGATCTGTGTGAGAGAGGCAAGATGACGCGCAAGTCTGG 1496  
Db 1423 ACTGTGTGACAGATGATGATCTGTGTGAGAGAGGGAAGATGACCGCCAAAGTCTGG 1482  
QY 1497 AGTCCGCAAGGCCATTTCTGGCGGCGAGCAAGGTGCGGCTGACCAAAAGTGAAGTCT 1556  
Db 1483 AGTCCGCAAGGCCATTTCTGGAGAGAGCAAGGTGCGGCTGACCAAAAGTGAAGTCT 1542

QY 1557 CCGCCAGATCGAACCCCAACCCCGGTGATCTGTCACCTCCAAACCAACATGTGCGCGTGA 1616  
Db 1543 CCGCCAGATAGACCCGACTCCCGGTGATCTGTCACCTCCAAACCAACATGTGCGCGTGA 1602  
QY 1617 TTGACGGGAACAGCACCCTTTCAGACCAACGAGCGGTTGACAGACCGGATGTTCAAT 1676  
Db 1603 TTGACGGGAACCTCAACGACCTTTCGAAACACGAGCGGTTGCAAGACCGGATGTTCAAT 1662  
QY 1677 TTGAATCAACCCCGCTGTGAGATGACTTTGGCAAGGTGACAAAGCAAGATCAAG 1736  
Db 1663 TTGAATCAACCCCGCTGTGATGATGACTTTGGGAAGGTCAACAGCAAGATCAAG 1722  
QY 1737 AGTTCTTCCGCTGGCGCAGAGATCAAGTGAACGAGTGGCGCATGAGTTCTACGTCAAG 1796  
Db 1723 ACTTTTCCGCTGGCGCAAGGATCAAGTGTGAGTGGAGCATGAATTTCTACGTCAAAA 1782  
QY 1797 AGGTGAGCCAAACAAAGACCCCGCCGATGACGCGGATTAACGAGCCCAAGCGG 1856  
Db 1783 AGGTGAGCCAAAGAAAGACCCCGCCAGTACGACGATATTAAGTGAAGCCCAACCGG 1842  
QY 1857 CCTGCCCTCAGTCCGGATCCATGACGTACAGCGCGGAAGAGCTCCGCTGACTTTG 1916  
Db 1843 TGCGCGAGTCAAGTTCGCGAGCCATGACGTACAGCGCGGA---AGTTGATCAACTACG 1899  
QY 1917 CCGACAGATCAACAAACAAATGTTCTGTCACGCGGCGATGCTTCAATGCTGTTCCCT 1976  
Db 1900 CAGACAGATCAACAAACAAATGTTCTGTCACGCGGCGATGATCTGATGCTGTTCCCT 1959  
QY 1977 GCAAGACATGCGAGAGAAATGAATCAAGATTTCAACATTTGCTTCAACGAGGAGAG 2036  
Db 1960 GCAGACATGCGAGAGAAATGAATCAAGATTTCAATATCTGCTTCACTCAACGAGAGAA 2019  
QY 2037 ACTGTTCAAGTGTCTTCCCGCGGTGACAAATCTCAACCGT---CGTCAGAAAGAGA 2093  
Db 2020 ACTGTTCAAGTGTCTTCC---CGTCAGAAATCTCAACCGT---CGTCAGAAAGAGA 2076  
QY 2094 CGTATCGAAACTCTGTGCAATGATCACTGTGCGGCGGCTCCGAGATGCTTGTCT 2153  
Db 2077 CGTATCGAAACTGTGCTATCATATCATATCATGGAAGAGTGCAGA---CGCTTGA 2133  
QY 2154 CGCGCTGCACTGTGTCACGTCGACCTGATGATGATGATGATGATGATGATGATGAT 2213  
Db 2134 CTGCTGCGATCTGTGTCATGATGATGATGATGATGATGATGATGATGATGATGAT 2193  
QY 2214 AAACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2273  
Db 2194 AAATCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2253  
QY 2274 GGCATTCGAGTGTGAGCTTGAACCTGAGACCCCGAAGCCCAAGCCCAACGAGCAA 2333  
Db 2254 GGAATTAAGACAGTGTGAGAGCTCAACCTGCGCCACCAACCAAGCCCGCAGAGCGG 2313  
QY 2334 AAGCAGACGACCGCGGCTGTGCTTCTGCTTCTGCTTCAAGTACCTCGACCTTCAAC 2393  
Db 2314 CATAGGAGACAGCAGAGGCTTGTGCTTCTGCTTCAAGTACCTCGACCTTCAAC 2373  
QY 2394 GGAATTCAGAGGGGAGCCCGTCAACGCGCGGAGCAGCAGCGGCTCGAGCAGCAAG 2453  
Db 2374 GGAATTCAGAGGGGAGCCCGTCAACGAGCAGAGCAGCGGCTCGAGCAGCAAA 2433  
QY 2454 GCCTACGACGACGCTCAAGCGGTCATCCGTAACCTGCGGTATTAACGAGCGGAC 2513  
Db 2434 GCCTACGACGAGCAGCTGACAGCGGAGACCAACCGTACCTCAAGTACCAACGAGCG 2493  
QY 2514 GCGAGTTTCAAGAGCGCTGCAAGAGATACGCTTTTGGGGGCAACCTCGGGGAGCA 2573  
Db 2494 GCGAGTTTCAAGAGCGCTTAAAGAGATACGCTTTTGGGGGCAACCTCGGAGAGCA 2553  
QY 2574 GTCTTCAAGCCAAAGAGCGGTTCTGAACTCTCGGTCTGTTGAGGAAGCGCTAAG 2633  
Db 2554 GTCTTCAAGCCAAAGAGAGGTTCTTGAACCTCTGCGCTGTTGAGGAACCTGTTAAG 2613

QY	2634	ACGGCTCCTGGAAAGAAACGTC	CCGGTAGAGCAGTCCGCCACAAGAC	CCAGACTCTCTCTCG	2693
Db	2614	ACGGCTCCGGAAAAAGAGG	CCGGTAGAGCAGACTCTCTGTGGAG	CCAGACTCTCTCG	2673
QY	2694	GGCATCCGCAGACAGCGCAGCAG	CCCGCTAAAAAGACTCAATTTTGT	CAGACTGGC	2753
Db	2674	GGAACCGGAAAGCGGGCCAGCAG	CTCGAAGAAAAAGATTGAATTTTGT	CAGACTGGA	2733
QY	2754	GACTCAGAGTCAGTCCCGGATTC	CAACCTCTCGAGAACTTCAGCAAC	CCCGCTGCT	2813
Db	2734	GACGCAGACTCAGTACCTGAC	CCCCAGCCTCTCGAGACGCCAC	CAGACGCCCTCTGCT	2793
QY	2814	GTGGACCTACTACAATGGCTTC	AGGCGGTGGCGCACCAATGGCAG	ACAATAACGAAGC	2873
Db	2794	CTGGGAATAATACGATGCTAC	AGGCAGTGGCGCACCAATGGCAG	ACAATAACGAAGC	2853
QY	2874	GCCGACGGAGTGGGTATGCTC	CAGGAAATTGGCATTGCGATTCC	ACATGGCTGGCGAC	2933
Db	2854	GCCGACGGAGTGGGTATGCTC	CGGAAATTGGCATTGCGATTCC	ACATGGATGGCGAC	2913
QY	2934	AGAGTCATCACCAACGACGAC	CCCGCCTTGCCCACTCAATAAC	CAACCTCTAC	2993
Db	2914	AGAGTCATCACCAACGACGAC	CCCGCCTTGCCCACTCAATAAC	CAACCTCTAC	2973
QY	2994	AAGCAATCTCCAGTCTTCAAC	GGGGCCAGCAAGCAACCACTAC	CTCGCTACAGC	3053
Db	2974	AAACAATTTCCAGCCAAATCA	---GGAGCCTCGAAGCAATCACT	ACTTTGGCTACAGC	3030
QY	3054	ACCCCTGGGGGTATTTTGAT	TTTCAACAGATTTCCACTTGCC	CACTTTCCACCCTG	3113
Db	3031	ACCCCTGGGGGTATTTTGAT	TTTCAACAGATTTCCACTTGCC	CACTTTCCACCCTG	3090
QY	3114	CAGCGACTCATCAACAACAAT	TGGGGATTCCGGCCCAAGACTC	CAACTTCAACTCTTC	3173
Db	3091	CAAGACTCATCAACAACAAT	TGGGGATTCCGACCCCAAGACTC	CAACTTCAACTCTTT	3150
QY	3174	AACATCCAAGTCAAGGAGTCA	CGACGAATGATGGCGTCACAAC	CACTGCTAATACTT	3233
Db	3151	AACATCCAAGTCAAGGAGTCA	CGCAGAATGACGTAAGCAGAC	GATGGCAATTAACCTT	3210
QY	3234	ACCAGCAGGTTCAAGTCTTCT	CGGACTCGAGTACCAAGTTC	CGTCTCGGCTCT	3293
Db	3211	ACCAGCAGGTTCAAGTCTTCT	CGGACTCGAGTACCAAGTTC	CGTCTCGGCTCG	3270
QY	3294	GCGCACCAGGCTGCTCCCTCC	GTTCGGTCCGCGGACGTTATG	ATTTCCGAATAACGC	3353
Db	3271	GCGCATCAAGGATGCTCCCG	CGCTTCCAGCAGACGCTTCT	CATGTGCCACAGTATGA	3330
QY	3354	TACCTGACGCTCAACAATGG	CAGACCCAAAGCCGTGGGAC	GTTTCACTTTTCTGCTGG	3413
Db	3331	TACCTGACGCTCAACAATGG	CAGACCCAAAGCCGTGTTCA	CTTTTCACTTTTCTGCTGG	3390
QY	3414	TATTTCCCTTCTCAGATGCT	GAGAACGGGCAACAATTTAC	CTTCACTACCTTTTGAG	3473
Db	3391	TACTTCTCTTCTCAGATGCT	GCCTACCGGAAACAATTTAC	CTTCACTACCTTTTGAG	3450
QY	3474	GAAATGCCCTTTCACAGAGCT	ACGCGCACAGCCAGGCTGGA	CCGCTGATGAATCCT	3533
Db	3451	GACGTTCCCTTTCACAGAGCT	ACGCTCACAGCCAGGCTGGA	CCGCTCATGAATCCT	3510
QY	3534	CTCATCGACCAATACCTGTAT	TACCTGAACAGAACTCAAAAT	CAGTCCGGAAGTGC	3593
Db	3511	CTCATCGACCAATACCTGTAT	TACCTGAACAGAACTCAAAAT	CAGTGAACCAACG	3570
QY	3594	AACAAGGACTTGCTGTTTAC	CCGCTGCTCCAGCTGGCATG	CTGTTCAAGCCAAAAC	3653
Db	3571	CAGTCAAGGCTTCAGTTTCT	CAGGCCGAGCGAGTGCATTC	GGGACCACTGTAAGAAC	3630
QY	3654	TGGCTAAGCTGAGCCCTGTAT	TCCGGCAGCAGCGCGTTCT	TAATAACAAAAACAGAC	3713
Db	3631	TGGCTTCTGAGCCCTGTAT	TCCGGCAGCAGCGAGTAT	CAAAAGACATCTGCGAT	3690
QY	3714	AACAGCAATTTTACCTGAG	CTGGTCTTCAAAATATAA	CTCAATGGCGGTGAAT	3773

Db	3691	AACAGTGAATAC	TGCTGGA	CTGGAG	CTACCA	GTACCA	GTACCA	CTCACT	CAATGG	CAGAG	ACTCT	GTG	3750															
QY	3774	ATCAACCC	TGGCA	CTGCTA	TGGCC	TCACA	CAACA	GACGA	AGCA	AGTTCT	TTCC	CATG	3833															
Db	3751	GTGAATC	CGGGCC	CGCCAT	GTGG	CAAG	CCAC	AGA	GCAT	GAAG	AAAGTT	TTTCT	CAG	3810														
QY	3834	AGCGGT	GCATG	ATTTT	TGGAAA	GAGAG	CGCGG	AGCTT	CAAA	CACTG	CATG	GACA	AT	3893														
Db	3811	AGCGGG	TTTCAT	CTTTG	GGAAG	CAAG	CTC	AGAAAA	CAAA	TGTGA	CACTT	GA	AAAG	3870														
QY	3894	GTCATG	ATTAC	AGAA	GAAATT	AAAG	CCATA	CCCTG	TGGC	ACCG	AAAG	ATT		3953														
Db	3871	GTCATG	ATTAC	AGAA	GAAATT	AAAG	CCATA	CCCTG	TGGC	ACCG	AAAG	ATT		3930														
QY	3954	GGGACC	TGGCA	GTCA	ATTTCC	AGAG	CA	GACAG	CA	CAACC	CTG	CGAC	CGGAG	ATGTCAT	4013													
Db	3931	GGTTCT	GTATC	TACCA	ACCTC	CAGAG	AGG	CA	CAAG	CA	GTAC	CCG	CAGAT	GTCAAC	3990													
QY	4014	GCTATG	GAGCA	TTA	CTGG	CA	TG	TGG	CA	AGAT	AGA	AGT	GTAC	CTG	CA	4073												
Db	3991	ACACA	GGCG	TTCTT	CCAG	CA	TG	GTG	CG	AGAG	CA	AGAT	GT	TACTT	CA	4050												
QY	4074	ATTTGG	CCAAA	TTCTC	AC	ACAG	ATG	CA	CTT	CA	CCCG	TCTC	CTT	AT	GG	CGC	4133											
Db	4051	ATCTGG	CAAG	ATTC	CA	CA	CG	AG	CA	TTT	CA	CCCT	CTC	CC	CT	CA	4110											
QY	4134	TTTGAC	TCA	GA	ACCCG	CTC	CAG	ATC	CT	CA	TCA	AAAA	CAC	GC	CTG	TCTG	CGAAT	4193										
Db	4111	TTGGA	CTTAA	CA	CCCTC	CTC	CA	CA	GA	TTT	CA	TCA	GA	GA	CA	CCCG	GTAC	CTGCGAAT	4170									
QY	4194	CCTCCG	GGGAG	TTT	CA	GCTA	CA	AA	AGTT	GT	CTT	CA	TTC	AT	CA	CCCA	AT	ACTT	CCA	CAG	GA	4253						
Db	4171	CCTCGA	CCAC	CTT	CAG	TG	CGG	CA	AA	AGTT	GT	CTT	CA	TCA	CA	CA	CA	AGT	ACTT	CCA	CAG	GA	4230					
QY	4254	CA-AGT	AGTGT	GAA	ATTGA	ATGG	GAG	CTG	CA	AA	AAAA	CA	GA	AA	AG	CG	CTG	GA	ATCC				4312					
Db	4231	CACG	GTAG	CGTG	AGAT	CGA	TGG	GAG	CTG	CA	AA	AAAA	CA	GA	AA	AG	CG	CTG	GA	ATCC			4290					
QY	4313	CGA	GTG	CAG	TAC	ATCC	AA	TATG	CA	AAAA	TTG	CA	AA	CGT	TG	ATT	T	ACT	GTG	CA	AA		4372					
Db	4291	CGAA	ATT	CAG	TAC	TTCC	AA	CTAC	AA	CA	AGT	CTG	T	AT	CGT	G	ACT	T	AC	CGT	G	AT	AC	4350				
QY	4373	CAATG	ACTT	T	A	CTG	AG	CC	TG	CG	CC	CA	T	TG	CA	CC	CG	T	T	AC	CC	CT	G	TA	4432			
Db	4351	TAATG	CGGT	TAT	T	CAG	AG	CC	TG	CG	CC	CA	T	TG	CA	CC	CA	G	AT	T	AC	CT	G	TA	4410			
QY	4433	ATTACG	TGTT	AT	CA	TAA	AC	CG	GT	G	ATT	CG	TTT	CA	G	T	G	A	CTT	T	G	GT	CT	G	TA	4492		
Db	4411	ATTG	CTGT	T	AT	CA	TAA	AC	CG	TT	AT	CG	T	G	CA	G	T	G	A	CTT	T	G	GT	CT	G	TA	4470	
QY	4493	TCTTAT	CTTAT	CG	T	T	AC	CA	TG	T	AT	AG	CTT	CA	CA	CTT	A	CTG	T	T	G	CG	CTT	G	C	4552		
Db	4471	TTCTT	CTTAT	CT	AG	TTT	C	ATG	CT	AG	T	AG	ATA	TT	A	G	AT	G	AT	G	CG	GT	T	AA	T	TA	4530	
QY	4553	GATA	AAAGA	CTT	AG	T	CAT	CG	GT	T	AC	CC	CT	AG	T	GA	GT	T	G	CC	CA	CT	CC	T	CT	G	C	4612
Db	4531	CTACA	AGG	-----	-----	-----	-----																					

RESULT 6  
US-09-770-315-3  
; Sequence 3, Application US/09770315  
; Patent No. 6429001  
; GENERAL INFORMATION:



APPLICANT: Chiron Corporation  
; TITLE OF INVENTION: Recombinant AAV Packaging Systems  
; FILE REFERENCE: 20263-501  
; CURRENT APPLICATION NUMBER: US/09/770,315  
; PRIOR FILING DATE: 2001-01-26  
; PRIOR APPLICATION NUMBER: US 60/178,536  
; PRIOR FILING DATE: 2000-01-26  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 3  
; LENGTH: 7557  
; TYPE: DNA  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: recombinant DNA  
US-09-770-315-3

Query Match 63.4%; Score 2993.4; DB 4; Length 7557;  
Best Local Similarity 80.4%; Pred. No. 0;  
Matches 3633; Conservative 0; Mismatches 841; Indels 45; Gaps 9;

QY 18 GCGCGCTCGCTCGGTGGGGCTCGGACCAAGGTCCGACAGCGGAGAGCTCTGC 77  
DB |||||  
4 GCGCGCTCGCTCGCTCACTGAGGCCCGCGCAAGCCCGGCGCTCGGCGACCTTTGG 63  
QY 78 TCTGCCCCCGCCACCGAGCGAGCGCGGAGAGAGAGAGAGAGAGAGAGAGAG 137  
DB |||||  
64 TCGCCCCCGCTCAGTAGCGAGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAG 123  
QY 138 GGGTAAT-----CGCGAAGCGCCTCCACGCTGCGCGCTC 172  
DB |||||  
124 TGGAGCTTGCTAGCATTAATTCTGATATGCTATACGAAGTTATCCGAGGGGTGG 183  
QY 173 AGCGCTGACGTAAATTACGTATAGG--GAGTGTCTGTATTAAGCTGTACGTAGT 229  
DB |||||  
184 AGTGTGACGTGAATTAAGTATAGGAGTCTGTATTAAGAGTCACTGAGT 243  
QY 230 GCTTTGCGACATTTTGGACACCAACGTCGTCATTTAGGATATATAGCCGAGT 289  
DB |||||  
244 G-TTTTGGACATTTTGGACACCAACGTCGTCATTTAGGATATTAAGCCGAGT 302  
QY 290 AGCAGATCTCATTTTG-ACCAGCAATTGGAACGACAGCAGCAGCCGCGCTCTA 348  
DB |||||  
303 CGCAGGCTCTCATTTTGAAGCGGAGGTTGAACGCGCAGCCGCGCATGCGGGTTTA 362  
QY 349 CGAGATCGTATCAAGGTGCCGAGCGACCTGAGCAGCAGCAGCAGCAGCAGCAG 408  
DB |||||  
363 CGAGATGTGATTAAGGTCCCGACGACCTTGAAGCATCTGCCGCGCATTTCTGAC 422  
QY 409 GTTTGTGAGTGGGTGCCGAGAGAAATGGAGCTGCCCCCGATTTCTGACATGATCT 468  
DB |||||  
423 CTTGTGAAGTGGGTGCCGAGAGAAATGGAGTTCGCGCAGATTTCTGACATGATCT 482  
QY 469 GAATCTGATGACGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 528  
DB |||||  
483 GAATCTGATGACGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 542  
QY 529 CCAATGCGCGCGTGAAGTAAAGGCCCGCGAGCCCTCTTTTGTTCAGTTCAGAGA 588  
DB |||||  
543 GGAATGCGCGCGTGAAGTAAAGGCCCGCGAGCCCTTTTGTTCAGTTCAGAGA 602  
QY 589 CGAGTCTTACTTCACTCCATATTTCTGGTGAAGACCAAGGGGTCAATTCATGCTGCT 648  
DB |||||  
603 AGAGAGCTACTTCCACATGACGCTGCTGGAAGAACACCGGGGTGAATTCATGCTT 662  
QY 649 GGGCGCTTCTGAGTCAAGTAAAGGAGCAGTGTGACAGCAGCAGCAGCAGCAGCAG 708  
DB |||||  
663 GGGAGCTTCTGAGTCAAGTAAAGGAGCAGTGTGACAGCAGCAGCAGCAGCAGCAG 722  
QY 709 GCGGACCTTCCCAACTGTTGCGGCTGACCAAGCAGCAGTAAATGCGCGGAGGGGAA 768  
DB |||||  
723 GCGGACCTTCCCAACTGTTGCGGCTGACCAAGCAGCAGTAAATGCGCGGAGGGGAA 782

QY 769 CAAGGTGTGACGAGTGTCTACATCCCACTACTCTCTGCCAAGACTCAGCCGAGCT 828  
DB |||||  
783 CAAGGTGTGATGAGTGTCTACATCCCACTACTCTCTGCCAAGACTCAGCCGAGCT 842  
QY 829 GCAGTGGCGGTGACTAATGAGAGATATATAAGCGCTTTTGAACCTGCGGAGCG 888  
DB |||||  
843 CCAGTGGCGGTGACTAATGAGAGATATATAAGCGCTTTTGAACCTGCGGAGCG 902  
QY 889 CAAACGCTGCTGGCGGACGACCTGACCCAGCTCAGCCAGACCCAGAGAGAGAGAG 948  
DB |||||  
903 TAAACGCTGCTGGCGGACGACCTGACCCAGCTCAGCCAGACCCAGAGAGAGAG 962  
QY 949 GAATCTGAACCCCAATTCTGACGCGCTGTCTATCCGCTCAAAAACCTCCGCGCTAC 1008  
DB |||||  
963 GAATCAGAAATCCCAATTCTGATGCGCGGTGATCAGATCAAAAACCTCAGCCAGTAC 1022  
QY 1009 GAGCTGTGCTGGGTGCTGTGGAACCGGGCATCACCCTCCGAGAAAGCAGTGCAGGA 1068  
DB |||||  
1023 GAGCTGTGCTGGGTGCTGTGGAACCGGGCATCACCCTCCGAGAAAGCAGTGCAGGA 1082  
QY 1069 GAGCAGGCTCTGATACATCTCTTCAACGCGCTTCCAACTCGCGGTCCAGATCAAGGC 1128  
DB |||||  
1083 GAGCAGGCTCTGATACATCTCTTCAATGCGCTTCCAACTCGCGGTCCAAATCAAGGC 1142  
QY 1129 CGCTTGAACATATGCGGCAAGATCATGCGCTGACCAAAATCCGCGCGCTACCTGT 1188  
DB |||||  
1143 TGCCTTGAACATATGCGGCAAGATCATGCGCTGACCAAAATCCGCGCGCTACCTGT 1202  
QY 1189 AGGCCCCCTCGCGCGCGGACATTAACCAACCGCATCTACCGCATCTGAGCTGAA 1248  
DB |||||  
1203 GGGCCAGCAGCCCGTGAGGACATTTCCAGCAATCGGATTTAATAATTGGAACCTAAA 1262  
QY 1249 CGGCTTGAACATATGCGGCAAGATCATGCGCTGACCAAAATCCGCGCGCTACCTGT 1308  
DB |||||  
1263 CGGCTTGAACATATGCGGCAAGATCATGCGCTGACCAAAATCCGCGCGCTACCTGT 1322  
QY 1309 GAAGCGCAACACCATCTGCTGTTGGCGCGGACCAACCGGCAAGACCAATCGCGGA 1368  
DB |||||  
1323 CAAGGAAACACCATCTGCTGTTGGCGCGGACCAACCGGCAAGACCAATCGCGGA 1382  
QY 1369 AGCCATCGCCACGCGCGCTCTTCAAGGCTGCTCAATGAGAACTTCC 1428  
DB |||||  
1383 GGCATAGCCACACTGTGCTTCAAGGCTGCTCAATGAGAACTTCC 1442  
QY 1429 CTTCAATGATTTGCTGACAGATGATCTGTTGGAGAGAGGCAAGATGACGCGCAA 1488  
DB |||||  
1443 CTTCAATGATTTGCTGACAGATGATCTGTTGGAGAGAGGCAAGATGACGCGCAA 1502  
QY 1489 GGTGCTGAGTCCGCGCAAGGCTTCTGCGCGGACAGAGTGCCTGGAACCAAAAGTG 1548  
DB |||||  
1503 GGTGCTGAGTCCGCGCAAGGCTTCTGCGCGGACAGAGTGCCTGGAACCAAAAGTG 1562  
QY 1549 CAACTGCTCGCCAGATGACCCCAACCCCGTGTGATGCTCACTCCAAACCAACATGTG 1608  
DB |||||  
1563 CAACTGCTCGCCAGATGACCCCAACCCCGTGTGATGCTCACTCCAAACCAACATGTG 1622  
QY 1609 CGCGTGAATGACGGAACAGCAACACCTTGAAGACAGCAGCAGCAGCAGCAGCAG 1668  
DB |||||  
1623 CGCGTGAATGACGGAACAGCAACACCTTGAAGACAGCAGCAGCAGCAGCAGCAG 1682  
QY 1669 GTTCAATTGAACTCAACCGCGCTGTGAGCATGCTTTGGCAAGGTGACAAAGCAGGA 1728  
DB |||||  
1683 GTTCAATTGAACTCAACCGCGCTGTGAGCATGCTTTGGCAAGGTGACAAAGCAGGA 1742  
QY 1729 AGTCAAGAAGTCTTCCGCTGGCGGACAGATCACTGACCGAGGTGCGCATGATTTCTA 1788  
DB |||||  
1743 AGTCAAGAAGTCTTCCGCTGGCGGACAGATCACTGACCGAGGTGCGCATGATTTCTA 1802  
QY 1789 CGTCAAGAAAGGTGAGCCAAACAAAGACCCCGCGGATGACCGGATTAAGCGAGCC 1848  
DB |||||  
1803 CGTCAAGAAAGGTGAGCCAAACAAAGACCCCGCGGATGACCGGATTAAGCGAGCC 1862  
QY 1849 CAAAGCGGCTGCGCTGAGTCCGCGATCATGACGCTGACAGCGGAGAGGCTCCGCT 1908



Db 1863 CAAACGGGTCGCGAGTCAAGTTCGCGAGCCATCGACGTACAGCGGGA--AGCTTCGAT 1919  
QY 1909 GGACTTTGCGCACAGGTACCCAAAACAATGTTCTCGTACCGCGGCATGCTTCAGATGCT 1968  
Db 1920 CAACTACGAGACAGGTACCAAAAACAATGTTCTCGTACCGTGGCATGAATCTGATGCT 1979  
QY 1969 GTTCCCTCGAAGACATCGAGAGATGAATCAGAAATTTCAACATTTGCTTCACGACGG 2028  
Db 1980 GTTCCCTCGAAGACATCGAGAGATGAATCAGAAATTTCAACATTTGCTTCACGACGG 2039  
QY 2029 GACGAGACTGTTAGAGTCTTCCCGGCGTGTCAAGATCTCAACCGGT--CGTCAG 2085  
Db 2040 ACAGAAAGACTGTTAGAGTCTTCC--CGTGCAGAACTCAACCGTTCTGTCGT 2096  
QY 2086 AAAGAGACGTATCGAAACTCTGTCCATTATCATCTGTGGGCGGGCTCCGAGAT 2145  
Db 2097 CAAAAGGGGTATCAGAACTGTGTACATTCATATCATGGAAGAAGTGCCAGA-- 2153  
QY 2146 TGCTTGCTCGGCTCGGATCTGTCTCAAGTGAACCTGATGATGCTGTTCTGAGCAATA 2205  
Db 2154 CGCTTGCACTGCTCGGATCTGTCTCAATGTGATTTGATGATGCTGATCTTTGAACAATA 2213  
QY 2206 AATGACTTTAAACAGGTATGCTGCTCCGATGTTATCTTCCAGATTTGCTCGAGACAACC 2265  
Db 2214 AATGATTTAAATCAGGTATGCTGCTCCGATGTTATCTTCCAGATTTGCTCGAGACACTC 2273  
QY 2266 TCTTGAGGGCATTCGCGAGTGTGGGACTTGAAACCTGAGCCCCGAAGCCAAAGCCA 2325  
Db 2274 TCTTGAGGAATAAGACAGTGTGAGACTCAAACTGCGCCACCAACCAAGGCCCG 2333  
QY 2326 ACCAGAAAAGCAGACGACGCGCGGCTGTGCTTCTGCTTACAAGTACCTCGGAC 2385  
Db 2334 CAGAGCGGCATAAGACGACGACGAGGGTCTGTGCTTCTGCTTACAAGTACCTCGGAC 2393  
QY 2386 CCTTCAACGCACTCGACAAAGGGGAGCCGCTCAACGCGCGGACGACGCGCCCTCGAGC 2445  
Db 2394 CCTTCAACGCACTCGACAAAGGGGAGCCGCTCAACGAGGACGACGCGCCCTCGAGC 2453  
QY 2446 ACACAAAGCCTTACGACGACGCTCAAAAGCGGGTGACAATCCGTACCTGCGTAAACC 2505  
Db 2454 ACACAAAGCCTTACGACGCGGAGCTGACAGCGGAGACAAACCGTACCTCAAGTAAACC 2513  
QY 2506 ACGCCGACGCGAGTTTCAAGAGCGTCTGCAAGAAGATACGTTTGGGGCAACCTCG 2565  
Db 2514 ACGCCGACGCGAGTTTCAAGAGCGCTTAAAGAAGATACGTTTGGGGCAACCTCG 2573  
QY 2566 GCGCAGAGTCTTCCAGGCGCAAGAACGGGTTCTCGAACCTCTCGGCTGTGAGGAAG 2625  
Db 2574 GACGAGAGTCTTCCAGGCGAAAGAGGGTCTTGAACTCTGGGCTGTGAGGAAG 2633  
QY 2626 GCGCTAAGACGGCTCTGGAAGAAAGCTCCGGTAGAGCAGTCCGCAACAGCCAGACT 2685  
Db 2634 CTGTTAAGACGGCTCCGGGAAAAAGAGCGCGGTAGAGCACTCTCTGTGAGCCAGACT 2693  
QY 2686 CCTCTCGGCACTCGGCAAGACAGGCGCAAGCCCGCTAAAGAAAGACTCAATTTGCTC 2745  
Db 2694 CCTCTCGGCAACCGGAAAGCGGGCGACAGCCTGCAAGAAAAAGATTGAATTTGCTC 2753  
QY 2746 AGACTGGGACTCAGAGTCACTCCCGATCCCAACCTCTCGGAGAACTTCAGCAACCC 2805  
Db 2754 AGACTGGGAGACGAGACTCAGTACCTGAGCCCGACCTCTCGGACAGCCACGACGCCC 2813  
QY 2806 CCGCTGCTGTGGGACTTACTACATGCTTCAAGCGGTGCGGCAACCAATGGCAGACAATA 2865  
Db 2814 CCGTGTCTGTGGAACTAATACGATGCTTCAAGCAGTGGCGCAACCAATGGCAGACAATA 2873  
QY 2866 ACAGAGGCGCGACGAGTGGTAATGCTCAAGAAATTGGCATTTGCATTCACATGGC 2925  
Db 2874 ACAGAGGCGCGACGAGTGGTAATCTCGGGAATTTGGCATTCACATGGC 2933  
QY 2926 TGGGCGACAGATCATCACCAAGCACCAGCCTGCGCTTGGCCCACTCAATAACC 2985

Db 2934 TGGGCGACAGAGTCATCACCAACGACACCCGAACCTGGGCTCCGACCTTACAACAACC 2993  
QY 2986 ACCCTTACAAGCAAAATCTCCAGTGTCTTCAACGCGGGCCAGCAACGACAACCACTTTCG 3045  
Db 2994 ACCCTTACAAGCAAAATTTCCAGCCAATCA--GGAGCCTCGAAGCAATCACTACTTGG 3050  
QY 3046 GCTACAGACACCCCTGGGGTATTTGATTTCAACAGATTTCCACTGCCACTTTTACACAC 3105  
Db 3051 GCTACAGACACCCCTGGGGTATTTGATTTCAACAGATTTCCACTGCCACTTTTACACAC 3110  
QY 3106 GTGACTGGCAGCGACTCATCAACAATTTGGGGATTCGGGCCCAAGAGACTCAACTTCA 3165  
Db 3111 GTGACTGGCAAGACTCATCAACAACACTGGGGATTCGACCBAAGAGACTCAACTTCA 3170  
QY 3166 AACTCTTCAACATCCAGTCAAGGAGTCAAGCAGATGATGGCTCAACAACCATGCTA 3225  
Db 3171 AGCTCTTAAACATTCAAGTCAAGAGGTCAAGCAGATGACGGTACGACGACGATGCGA 3230  
QY 3226 ATAACTTACAGCAGCGTTCAAGTCTTCTCGACTCGGATACAGCTTCCGTAAGTCC 3285  
Db 3231 ATAACTTACAGCAGCGTTCAAGTCTTCTCGACTCGGATACAGCTTCCGTAAGTCC 3290  
QY 3286 TCGGCTTGCGCAGCAGCGGCTGCTCCCTCCGTTCCGCGGAGCGTTCATGATTCGCGC 3345  
Db 3291 TCGGCTTGCGCAGCAGCGGCTGCTCCCTCCGTTCCGCGGAGCGTTCATGATTCGCGC 3350  
QY 3346 AATACGGCTACCTGACCGCTCAACAATGGCAGCCAAAGCCGTGGGAGCTTCATCTTTACT 3405  
Db 3351 AGTATGATTAACCTCAACCTGAAACAAGGAGTCAAGGACGTAAGGACGCTCTTCAATTTACT 3410  
QY 3406 GCTTGAATATTTCCCTTCTCAGATGCTGAGAACGGGCAACAATTACCTTCACTACA 3465  
Db 3411 GCTTGAATATTTCCCTTCTCAGATGCTGAGAACGGGCAACAATTACCTTCACTACA 3470  
QY 3466 CTTTGAAGGAGTCCCTTTCACAGCAGCTACGCGCAGCAGAGCCGTGAGCCGGCTGA 3525  
Db 3471 CTTTGAAGGAGTCCCTTTCACAGCAGCTACGCTCAAGCCAGAGTGTGAGCCGGCTGA 3530  
QY 3526 TGAATCTCTCATGACCAATTAACCTGTATTACCTGAACAGAACTCAAAATCAGTCCGAA 3585  
Db 3531 TGAATCTCTCATGACCAATTAACCTGTATTACCTGAACAGAACTCAAAATCAGTCCGAA 3590  
QY 3586 GTGCCAAAACAAGGACTGTGCTTTAGCCGCTGGGCTCCAGCTGGCATGTCTGTTACG 3645  
Db 3591 CCACCAACGCAAGTCAAGGCTTCAAGTTTCTCAAGCCGAGCGAGTGAATTCGGGACCACT 3650  
QY 3646 CCAAAAACCTGGTACCTGACCCCTGTATTCGGCAGACGCGGCTTCTTAAACAACAAAACAG 3705  
Db 3651 CTAAGAACTGGCTTCTGAGCCCTGTATACCGCAGCAGAGTATCAAGACATCTGCGG 3710  
QY 3706 ACAACAACAACAGCAATTTTACCTGAGCTGTGCTTCAAAATATTAACCTCAATGGGCGTG 3765  
Db 3711 ATAAACAACAAGTGAATACCTGTGAGCTGAGCTACCAAGTACCACTCAATGGCAGAG 3770  
QY 3766 AATTCATCATCAACCTGGCACTGCTATGCGCTCAACAAGAAGCAGAGCAAGCAAGTCT 3825  
Db 3771 ACTCTGTGTAATCCGGGCCCGGCATGGAAGCACAAGACGATGAAGAAAAGTTT 3830  
QY 3826 TTCCATGAGCGGTGTATGATTTTGGAAAAAGAGCGCCGAGCTTCAACAACATGCAAT 3885  
Db 3831 TTCTCAGAGCGGGGTTCATCTTTGGGAGCAGAGCTCAGAGAAAACAAATGTGAGCA 3890  
QY 3886 TGAACAATGTATGATTAACAGCAAGAGGAAATTAAGCCAATAACCTGTGCGCACCG 3945  
Db 3891 TTGAAAAGGTATGATTAACAGCAAGAGGAAATCAGGACAACCAATCCCGTGGCTAACGG 3950  
QY 3946 AAAGATTTGGGACCGTGGAGTCAATTTCCAGAGCAGCAGACAGACCTGCGACCGGAG 4005  
Db 3951 AGCAGTATGTTCTGTATTAACCAACTCCAGAGGCGAACAGACAAGCAGTACCGCAG 4010  
QY 4006 ATGTGATGCTATGAGGACTTACCTGGCATGTGTGCAAGATAGAGACGTTGACTTGC 4065  
Db 4011 ATGTCAACACACAAGGCGTTCTTCCAGGACATGTTGCGCAGACAGAGATGTACTTTC 4070

QY	4066	AGGGTCCCATTTGGGGCCAAAATTCTCTCACACAGATGGACACTTTCACCCGCTCTCTTA	4125
Db	4071	AGGGGGCCCATCTGGGCAAGATTTCACACACGAGCGACATTTCACCCTCTCCCTTCA	4130
QY	4126	TGGGGCGGCTTTGGACTCAAGAACCCGCTCTCAGATCCTCATCAAAAAACAGCGCTGTTC	4185
Db	4131	TGGGTGGATTGGACTTAAACACCCTCTCTCCACAGATTCTCATCAAGAACACCCCGGTAC	4190
QY	4186	CTGCGAATCCTCCGGCGGAGTTTTCAGCTACAAAGTTTGCTTCATTCATCACCCCAATACT	4245
Db	4191	CTGCGAATCCTTCAGACCACCTTCAGTGGCGCAAAAGTTTGCTTCTTCATCACACAGTACT	4250
QY	4246	CCACAGGACAAAGTGAAGTGTGGAATTGAAATGGAGCTGCAGAAAAGAAAAACAGCAAGCGCT	4305
Db	4251	CCACGGGACAGGTCAAGCTGGAGATCGAGTGGAGCTGCAGAAAGAAAAACAGCAAAAGCT	4310
QY	4306	GGAATCCCCGAAGTGCAGTACATCCAAATTATGCAAAATCTGCCAAGCTTGATTTTACTG	4365
Db	4311	GGAATCCCCGAATTCAGTACACTTCCAACATAACAACAAGTCTGTTAATGTGACTTTACTG	4370
QY	4366	TGGACACAATGGACTTTATACTGAGCCTCGCCCATTTGGCACCCGTTACCTTACCCTGTC	4425
Db	4371	TGGACACTAATGCGCTGTATTCAGAGCCTCGCCCATTTGGCAACAGATACTGACTCGTA	4430
QY	4426	CCCTGTATTAATGCTGTTAATCAATAAACCGGTTGATTCGTTCAAGTTGAACCTTGGTCTC	4485
Db	4431	ATCTGTAAATTGCTTGTTAATCAATAAACCGTTTAAATTGCTTCAAGTTGAACCTTGGTCTC	4490
QY	4486	CTGTCCTTCTTATCTTATC 4504	
Db	4491	TGCGTAATTTCTTCTTATC 4509	

```

RESULT 7
US-09-532-594B-1
; Sequence 1, Application US/09532594B
; Patent No. 6468524
; GENERAL INFORMATION:
; APPLICANT: Chorini, John A.
; APPLICANT: Kotin, Robert M.
; APPLICANT: Safer, Brian
; APPLICANT: Davidson, Beverly
; TITLE OF INVENTION: AAV4 VECTOR AND USES THEREOF
; FILE REFERENCE: 14014.0252U2
; CURRENT APPLICATION NUMBER: US/09/532,594B
; CURRENT FILING DATE: 2000-03-22
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 4767
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence; No. 6468524e =
; NAME/KEY: misc_feature
; LOCATION: 3009
; OTHER INFORMATION: n = g, a, c or t(u)
; NAME/KEY: misc_feature
; OTHER INFORMATION: AAV4 genome
;
US-09-532-594B-1

```

Query Match	55.0%;	Score 2596;	DB 4;	Length 4767;
Best Local Similarity	74.5%;	Pred. No. 0;		
Matches 3582;	Conservative	0;	Mismatches 1096;	Indels 129;
				Gaps 20;

[illegible]

Db	61	AGACTCGCGCCTCTGGCCGGCAGGGCCGAGTGAGTGAACGACGCCGATAGAGGAGTTC	120
QY	121	GGAACCTCCATCACTAAGG-----GTATCGCGAAGCGCTCCAC	161
Db	121	GCCAACCTCCATCATTAAGTTTCCCACTGACGTCAATGTGACGTCTAGGGTTAGGAG	180
QY	162	GCTGCCGCGTCAGCGCTGACGTAATTAAGTCAATGAGGAGTGTCTGTA-----T	213
Db	181	GTCCCTGTATTAGCACTGACGTGAGTGTCTATTTCGGGAGCGTAGCGAGCGCATACC	240
QY	214	TAGCTGTACGTGA-----GTGCTTTTGGCAATTTTGGACACACACGTGGCCAT	263
Db	241	AAGCTGCCACGTCAACAGCACGTGTCCTGTTGGCAAGTTTGCACACCATGTGTGAG	300
QY	264	TTAGGGTATATATGGCCGACGTGAGCGAGC-AGGATCTCCATTTTGACCGCGAATTTGAA	322
Db	301	G-AGGGTATATTAACCGCGACGTGAGCCAGCGAGAGCTTCATTTTGGCCGAATTTGAA	359
QY	323	CGAGCAGCAGCCATGCCGGGCTTCTACGAGATCTGTATCAAGGTGCCGAGCGACCTTGAC	382
Db	360	CGAGCAGCAGCCATGCCGGGCTTCTACGAGATCTGTATCAAGGTGCCGAGCGACCTTGAC	419
QY	383	GAGCACCTGCCGGCATTTCTGACTCGTTTGTGAGCTGGGTGCCGAGAAAGAAATGGAG	442
Db	420	GAGCACCTGCCGGCATTTCTGACTCTTTTGTGAGCTGGGTGCCGAGAAAGAAATGGAG	479
QY	443	CTGCCCCCGGATTTCTGACATGATCTGAATCTGAATTGACAGCACCCCTGACCGTGCC	502
Db	480	CTGCCGCCGGATTTCTGACATGACTTGAATCTGAATTGACAGCACCCCTGACCGTGCC	539
QY	503	GAGAAGCTGCAGCGGCACTTCTGTGTCATGCGCGCGGTGAGTAAGCCCGGAGGCC	562
Db	540	GAAAAGCTGCACCGCGAGTCTCTGTGAGTGGCGCGGTGAGTAAGCCCGGAGGCC	599
QY	563	CTCTTCTTGTTCAGTTGAGAAAGGCGGAGTCTCTTCCACTTCATATTCTGTGAG	622
Db	600	CTCTTCTTGTTCAGTTGAGAAAGGCGGACAGCTACTTCCACTGACATCTGTGTGAG	659
QY	623	ACCACGGGGGTCAATTCATGTGTCTGGGCGGCTTCTGAGTCAGATTAGGACAAGCTG	682
Db	660	ACCGTGGCGTCAATTCATGTGTGTGGGCGGCTACTGAGCCAGATTAAAGAGAAGCTG	719
QY	683	GTGACAGCATCTACCGCGGGATCGAGCCGACCCCTGCCAACTGTTCCGGTGACCAAG	742
Db	720	GTGACCCGCACTACCGCGGGGTGAGCCGACCTTCCGAATCTGTTCCGGTGACCAAG	779
QY	743	ACCGGTATATGCGCGCGAGGGGGAACAAGTGTGTGACGAGTGCTACATCCCCAACTAC	802
Db	780	ACCGGTATATGCGCGCGAGGGGGAACAAGTGTGTGACGAGTGCTACATCCCCAACTAC	839
QY	803	CTCCTGCCCAAGACTCAGCCCGAGCTGCAGTGGGCGTGACTAACATGAGAGTATATA	862
Db	840	CTGCTCCCCAAGACCCAGCCCGAGCTCCAGTGGGCGTGACTAACATGAGCCAGTATATA	899
QY	863	AGCGCCTGTTTGAACCTGCGCGGAGCGCAACCGCTCGTGGCGGACGACCTGACCCACGT	922
Db	900	AGCGCCTGTTTGAATCTCGCGGAGCGTAAACCGCTGTGGCGGACGACATCTGACGACGTG	959
QY	923	AGCCAGACCCAGAGAGCAACAAGAGAATCTGAACCCCAATTCTGACGCGCTGTATC	982
Db	960	TCGCAGACGACGAGAGCAACAAGAAAAACAGAAACCCCAATTCTGACGCGCGGTATC	1019
QY	983	CGGTCAAAAACCTCCGCGGCTACATGAGCTGGTGGGTGGCTGTGACCGGGGCAATC	1042
Db	1020	AGGTCAAAAACCTCCGCGGCTACATGAGCTGGTGGGTGGCTGTGACCGGGGCAATC	1079
QY	1043	ACCTCCGAGAGCAATGATCCAGAGGACCAAGGCTTCGTACATCTCCTTCAACGCGCT	1102
Db	1080	ACGTCAAAAAAGCAATGATTCAGAGAGGACCAAGGCTTCGTACATCTCCTTCAACGCGCC	1139
QY	1103	TCCAACCTCGCGGTCCAGATCAAGGCGCTCTTGACAATGCCGCAAGATCATGGCGCTG	1162
Db	1140	TCCAACCTCGCGGTCACTAATCAAGGCGCGCTGACATGCTCCAAAATCATGAGCTG	1199







```
Db 3324 TTTTCCCAAGACGCTTTATGTGCCCCAGTAGCGCTACTGTGACTGTGACCGGCAA 3383
QY 3368 CAATGGCAGCCAAAGCCGTGGAGCTTCATCTTTTACTGCGCTGGAATATTTCCCTCTCA 3427
Db 3384 CACTTCGACGAAACAGACTGACAGAAATGCTTCTACTGCGCTGAGTACTTTCCTCGCA 3443
QY 3428 GATGCTGAGAACGGGCAACAACCTTTACCTTACGCTACACCTTTGAGGAAGTGCCTTCCA 3487
Db 3444 GATGCTGCGAGCTGGCAACAACCTTTGAATTAAGTACAGTTTGAGAGGTGCCTTCCA 3503
QY 3488 CAGCAGCTACCGGCAAGCCAGAGCCGTGACCGGCTGATGAATCTCTCATCGACCAATA 3547
Db 3504 CTGATGTACCGCAAGCCAGAGCCGTGACCGGCTGATGAACCTCTCATCGACCAATA 3563
QY 3548 CCTGTATTACTGAACAGAACTCAAAATCAGTCCGGA--AGTGCCTCAAAACAAGACTT 3604
Db 3564 CCTGTGGGAGCTGCAATCGAACCAACCGGAACCACTGATGCGGGACTGCCACAC 3623
QY 3605 GCTGTTAGCCGTGGTCTCCAGCTGGCATGTCTGTTCAGCCCAAAAAGTGGTACCTGG 3664
Db 3624 CAACTTACCAAGCTGCGGCTACCACTTTTCCAACCTTAATAAAGAACTGGCTGCCCG 3683
QY 3665 ACCCTGTTATCGGCAAGCGCGTCTTAACAACAAAACAGACACACACACCAATTT 3724
Db 3684 GCCTTCAATCAAGCAGCAGGGCTTCTCAAGACTGCCAATCAAACTACAGATCCCTGC 3743
QY 3725 TACCTGACTGTGCTTCAAAATATAAC-----CTCAATGGGCGTGAATC 3769
Db 3744 CACCGGCTCAGACAGTCTCATCAAAATACGACGACACGACACTCTGACGGAAGATGAG 3803
QY 3770 CATCATCAACCTGGCACTGCTATGCGCTCAACAAGACGACGAGCAAGTCTTCC 3829
Db 3804 TGCCCTGACCCCGGACCTTCAATGCGCACGCGTGAACCTGCGACAGCAAGTTCCAG-- 3860
QY 3830 CATGACGGTGTATGATTTTGGAAAAAGAGCGCGGAGCTTCAACAACACTGCATTGA 3889
Db 3861 CAACAGCAGCTCATCTTTGCGGGCTTAACAGAACGCAACACGCGCACCGTACC 3920
QY 3890 CAATGTCATGATTAACAGACGAAAGAAATTAAGCCACTAACCTGTGGCCACCGAAG 3949
Db 3921 GACTCTGATCTTACCTCTGAGGAGAGCTGGAGCCCAACCAACCGCCAGTACGGACAT 3980
QY 3950 ATTTGGGACCGTGGCAGTCAATTTCCAGAGCAGACACAGACCTGCGACCGGAGATGT 4009
Db 3981 GTGGGGCAACTTACTTGGCGGTGACCAAGACACCACTGCGGACCGTGGACAGACT 4040
QY 4010 GCATGCTATGAGCATTAACCTGGCATGTGTGGCAAGATAGAGACGTGTAACCTGAGGG 4069
Db 4041 GACAGCCTTGGAGCCGTGCTGGAATGTGTGGCAAAAACAGACATTTTACTACAGGG 4100
QY 4070 TCCCATTTGGGCAAAATTCCTCACAACAGATGACATTTCAACCGCTTCTTATGGG 4129
Db 4101 TCCCATTTGGGCAAGATTCCTCATACCGATGACATTTCAACCGCTGATGG 4160
QY 4130 CGGCTTTGACTCAAGAACCCGCTCCTCAGATCCTCATCAAAAACAGCGCTGTTCTGC 4189
Db 4161 TGGTGTGGGCTGAACAACCGGCTCCTCAAAATTTTATCAAGAACACCGGTAACCTGC 4220
QY 4190 GAATCCTCCGGCGAGTTTTCAGCTACAAGTTTGTCTCATTCATCAACCAATACTCCAC 4249
Db 4221 GAATCCTGCAACGACCTTCAAGCTTACTCCGGTAAACTCCTTCATTACTAGTACAGCAC 4280
QY 4250 AGGACAAGTGAAGTGAAGTGAATGGAGCTGCAAGAAAGAAACAGCAAGCGCTGGA 4309
Db 4281 TGGCCAGGTGTGGTGAAGTGAATGAGTGGAGATCCAGAAAGAGCGGTCCAAACGCTGGA 4340
QY 4310 TCCCGAAGTCAAGTACATCCAAATTAAGCAAAATCTGCCAACGTTGATTTTACTGTGA 4369
Db 4341 CCCCAGGTCAAGTTTAACTCCAACTACGAGACGCAAAACTCTGTGTGGGCTCCCGA 4400
QY 4370 CAACAATGACTTATTAAGCTGAGCCTGCGCCCAATTGGCAACCGTACTTACCCGCT 4429
```

```
Db 4401 TGGCGTGGGAAATACACTGAGCCTAGGGCTATCGGTACCCGCTAACCACCAACCACT 4460
QY 4430 GTAATTACGTGTTAATCAATAAACCGGTGATTCGTTTCAAGTTGAACCTTGGTCTCC-TG 4488
Db 4461 GTAATAACCTGTTAATCAATAAACCGGTTAATTCGTTTCAAGTTGAACCTTGGTCTCCGTG 4520
QY 4489 TCCCTTATCTTATC-GTTACCATGGTTATAGCTTACATTAACCTG-----CTTG 4541
Db 4521 TCCCTTATCTTATCTGTTTCATAGGCTACTGCGTACATTAAGCAGCGGCTGCGCGC 4580
QY 4542 TTGGCGCTTGGGATAAAGACTTACG-----TCATCGGGTTACCCCTAGTGAATGG 4591
Db 4581 TTGGCGCTTGGGCTTACACTGCGGTTAATCAGTAACTTCTGGCAAAACAGATGATGG 4640
QY 4592 AGTTGCCCACTCCCTCTGCGCGCTCGCTCGCTCGGTGGGCGCTGCGCAAAAGTCC 4651
Db 4641 AGTTGGCCACATTAAGCTATGCGCGCTCGCTCACTCACTCGGCGCTGGAACCAAGTCT 4700
QY 4652 GCAGACGCGCAGAGCTGCTCTGCGCGGCCCCACCGACGCGAGCGGCGCAGAGGGAG 4711
Db 4701 CCAGACTGCGCGCTCTGCGCGGCGGCGGCGAGTGAAGTGAAGCGCGCATAGAGGAG 4760
QY 4712 TGGGCAA 4718
Db 4761 TGGCCAA 4767
```

```
RESULT 8
US-09-770-315-4
; Sequence 4, Application US/09770315
; Patent No. 6429001
; GENERAL INFORMATION:
; APPLICANT: Chiron Corporation
; TITLE OF INVENTION: Recombinant AAV Packaging Systems
; FILE REFERENCE: 20263-501
; CURRENT APPLICATION NUMBER: US/09/770,315
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: US 60/178,536
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 4072
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: recombinant DNA
US-09-770-315-4
```

```
Query Match 35.8%; Score 1691.2; DB 4; Length 4072;
Best Local Similarity 79.1%; Pred. No. 0;
Matches 2063; Conservative 0; Mismatches 533; Indels 12; Gaps 4;

QY 1900 AGTCCGCTGACTTTGCCGACAGGTACCAAAACAATGTTCTGTCACGCGGCATGCT 1959
Db 1164 AGCTTGATCACTACGACAGAGGTACCAAAACAATGTTCTGTCACGTCGATGAA 1223
QY 1960 TCAGATGCTGTTTCCCTGCAAGACATGCGAGAGATGAATGAATTTCAACATTTGCTT 2019
Db 1224 TCTGATGCTGTTTCCCTGCAAGACATGCGAGAGATGAATGAATTTCAATATCTGCTT 1283
QY 2020 CACGACGGGACGAGAGACTGTTCAAGAGTCTTCCCGCGGTGTCAGAACTCAACCGGT 2079
Db 1284 CACTCACGGAACGAAAGACTGTTTGAAGTCTTCC--CGTGTCAGAATCTCAACCCGT 1340
QY 2080 --CGTGAAGAGAGAGCGTATGGAATACTGTGCCATTCATCATCTGCGGCGGCG 2136
Db 1341 TTCTGTCTCAAAAGGCGTATCAGAAACTGTGCTACATTCATATCATATGGAAGGT 1400
QY 2137 TCCCGAGATGCTTGTGCGGCTGCGATCTGTGTCACGTCGATGATGATGATGATGAT 2196
Db 1401 GCCAGA---CGTTGCACTGCTGCGATCTGTGTCATGATGATGATGATGATGATGAT 1457
```



Db 3615 ACTTACTGTGACACTAATGCGGTATATCAAGCCCTGCCCATTTGGACACAGATACC 3674  
QY 4417 TTACCCGTCCTCCCTGAATTAAGTGTATTAATCAATAAACCGGTGATGTTGAGTTGAGTGAAC 4476  
Db 3675 TGAATCGTATCTGTAATGCTTGTATTAATCAATAAACCGTTTATTCGTTCAAGTTGAAC 3734  
QY 4477 TTTGGTCTCTGTCCTTCTTATCTATC 4504  
Db 3735 TTTGGTCTCTGTCCTTCTTCTTATC 3762

RESULT 9  
US-09-532-594B-3

Sequence 3, Application US/09532594B  
Patent No. 6468524  
GENERAL INFORMATION:  
APPLICANT: Chordini, John A.  
APPLICANT: Kolin, Robert M.  
APPLICANT: Safer, Brian  
APPLICANT: Davidson, Beverly  
TITLE OF INVENTION: AAV4 VECTOR AND USES THEREOF  
FILE REFERENCE: 14014.025202  
CURRENT APPLICATION NUMBER: US/09/532,594B  
CURRENT FILING DATE: 2000-03-22  
NUMBER OF SEQ ID NOS: 22  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 3  
LENGTH: 1872  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence; No. 6468524e =  
OTHER INFORMATION: synthetic construct  
NAME/KEY: CDS  
LOCATION: (1)..(1872)  
NAME/KEY: misc feature  
OTHER INFORMATION: AAV4 Rep gene (full length)  
US-09-532-594B-3

Query Match 30.6%; Score 1444.8; DB 4; Length 1872;  
Best Local Similarity 86.4%; Pred. No. 0;  
Matches 1620; Conservative 0; Mismatches 249; Indels 6; Gaps 2;

QY 335 ATGCCGGGCTTCTACGAGATCGTATCAAGGTGCGGAGCGACCTGGACGACCTGCGG 394  
Db 1 ATGCCGGGCTTCTACGAGATCGTATCAAGGTGCGGAGCGACCTGGACGACCTGCGG 60  
QY 395 GGCATTCTGACTCGTTTGTGAGCTGGGTGGCGGAGAAAGATGGAGCTGCGCGGAT 454  
Db 61 GGCATTCTGACTCGTTTGTGAGCTGGGTGGCGGAGAAAGATGGAGCTGCGCGGAT 120  
QY 455 TCTGACATGATCTGAATCTGATTGAGCAGGACACCTGACCGTGGCGGAGAGCTGCAG 514  
Db 121 TCTGACATGATCTGAATCTGATTGAGCAGGACACCTGACCGTGGCGGAGAGCTGCAG 180  
QY 515 CGCGACTTCCTGTCATGCGCGCGCGCTGAGTAAGCGCCCGGAGCGCTTCTTTGTT 574  
Db 181 CGCGAGTTCCTGTCATGCGCGCGCGCTGAGTAAGCGCCCGGAGCGCTTCTTTGTT 240  
QY 575 CAGTTGAGAAAGGCGAGCTCTACTTCCACCTCCATATTTCTGTGAGACCAACGCGGCTC 634  
Db 241 CAGTTGAGAAAGGCGAGCTCTACTTCCACCTCCATATTTCTGTGAGACCAACGCGGCTC 300  
QY 635 AAATCCATGCTGTGGCGCGCTTCTGAGTCAGATTAGGGAACAAGCTGTGACAGCATC 694  
Db 301 AAATCCATGCTGTGGCGCGCTTCTGAGTCAGATTAGGGAACAAGCTGTGACAGCATC 360  
QY 695 TACCGCGGATCGAGCCGACCTGCGCAACTGTTGCGGCTGACCAAGACGCGTAATGGC 754  
Db 361 TACCGCGGATCGAGCCGACCTTCCGAAGTGTGCGGCTGACCAAGACGCGTAATGGC 420  
QY 755 GCCGAGGCGGAAAGGTGTGAGCAGAGTGTACATCCCAACTACCTCTGCGCAAG 814

Db 421 GCCGAGCGGGAAACAAGTGTGAGCAAGTGTACATCCCACTACCTGCTCCCAAG 480  
QY 815 ACTCAGCCGAGCTGCAGTGGCGGTGACTAATGAGAGATATATAAGCGCTGTTG 874  
Db 481 ACCCAGCCGAGCTGCAGTGGCGGTGACTAATGAGAGATATATAAGCGCTGTTG 540  
QY 875 AACCTGGCGAGCGCAAAAGGCTGGCGCAGCAGCAGTACCGACGTCAGCAGACCCAG 934  
Db 541 AATCTCGCGAGCGTAAACGGCTGGTGGCGCAGCATCTGACGACGTCGACAGACGAG 600  
QY 935 GAGCAGAACAGAGAAATCTGAACCCCAATTTCTGACGCGCTGTATCCGGTCAAAAACC 994  
Db 601 GAGCAGAACAGAGAAACCAAGACCCCAATTTCTGACGCGCGGTATCAGGTCAAAAACC 660  
QY 995 TCCGCGGCTACATGAGCTGTCGGGTGGCTGTGACCGGGGATCACCTCCGAGAG 1054  
Db 661 TCCGCGGCTACATGAGCTGTCGGGTGGCTGTGACCGGGGATCACCTCAGAAAAG 720  
QY 1055 CAGTGATCCAGAGAGACCGCTGCTACATCTCTTCAACGCGCTTCCAACTCCGG 1114  
Db 721 CAATGATCCAGAGAGACCGCTGCTACATCTCTTCAACGCGCTTCCAACTCCGG 780  
QY 1115 TCCAGATCAAGCGCGCTGTGACATGCGGCAAGATCATGCGCTGACCAATCCGG 1174  
Db 781 TCACAAATCAAGCGCGCTGTGACATGCTTCCAAATCATGAGCTGACAAAGACGGCT 840  
QY 1175 CCGCATCTGTGTAGCCCGCTGCGCGCGGACATTAACAACCGCATCTACCGC 1234  
Db 841 CCGCATCTGTGTAGCCCGCTGCGCGCGGACATTAACAACCGCATCTACCGC 900  
QY 1235 ATCTGAGCTGAACGGCTACGAACCTGCTACGCGCGCTCGTCTTCTCGGTGGCG 1294  
Db 901 ATCTGAGCTGAACGGGTACGATCCGACAGTACGCGCGCTCGTCTTCTCGGTGGCG 960  
QY 1295 CAGAAAGGTTCCGGGAAGCGCAACCATCTGCTGTTGGCGCGGCCACACCGGCAAG 1354  
Db 961 CAAAGAGTTCCGGGAAGGAGAACCATCTGCTCTTTGGCGCGGCCACACCGGTAAG 1020  
QY 1355 ACCAATCGCGGAAAGCATCGCCGCGCGCTGCTTCTACGCGCTCGTCACTGAGC 1414  
Db 1021 ACCAATCGCGGAAAGCATCGCCGCGCGCTGCTTCTACGCGCTCGTCACTGAGC 1080  
QY 1415 AATGAACTTCCCTCAATGATGCTGACCAAGATGATGATGTTGGGAGAGAGGC 1474  
Db 1081 AATGAACTTCCCTCAATGATGCTGACCAAGATGATGATGTTGGGAGAGAGGC 1140  
QY 1475 AAGATGAGCGCCAAAGTGTGAGTCCGCAAGGCCATTTCTGCGCGGACGAAGGTGCGC 1534  
Db 1141 AAGATGAGCGCCAAAGTGTGAGTCCGCAAGGCCATTTCTGCGCGGACGAAGGTGCGC 1200  
QY 1535 GTGACCAAAAGTCAAGTGTGCGCGCCAGATCGACCCCAACCCCGTATCGTCACTCC 1594  
Db 1201 GTGACCAAAAGTCAAGTGTGCGCGCCAGATCGACCCCAACCCCGTATCGTCACTCC 1260  
QY 1595 AACACCAATGTGCGCGCTGATTTGACCGGAAACAGACCACTTTCAGACACAGCAGCGC 1654  
Db 1261 AACACCAATGTGCGCGCTGATTTGACCGGAAACAGACCACTTTCAGACACAGCAGCGC 1320  
QY 1655 TTGACGAGCGGATGTTCAATTTGAATCAACCGCGCTGTGACATGACTTTGGCAAG 1714  
Db 1321 CTCCAGAGCGGATGTTCAATTTGAATCAACCGCGCTGTGACATGACTTTGGCAAG 1380  
QY 1715 GTGACCAAGCAAGATCAAAAGTCTTCCGCTGGCGCAGAGTCAAGTCAAGCGAGGTG 1774  
Db 1381 GTGACCAAGCAAGATCAAAAGTCTTCCGCTGGCGCAGAGTCAAGTCAAGCGAGGTG 1440  
QY 1775 GCGCATGATTTCAAGTCAAGAAAGGTGAGCCAAACAAAGACCGCGCCGATGAGCGG 1834  
Db 1441 ACTCAGAGTTTACGTACAGAAAGGTGAGCTAAGAAAGAGCGCGCCCAATGACGCA 1500  
QY 1835 GATAAAGCAAGCCCAAGCGGCTGCGCTCAGTGCAGATCAATGACGTCAGACGCG 1894  
Db 1501 GATATAAGTACCCCAAGCGGCTGTCTCAGTGTGCGCAGCATCGACGTACAGACGCG 1560



QY 1895 GAAGAGCTCCGGTGACTTTCGCCAGAGTACCAGAAACAATGTTCTCGTACGCGGC 1954  
 Db 1561 GA--AGCTCCGGTGACTACGCGAGAGTACCAAAAACAATGTTCTCGTACGCGGT 1617  
 QY 1955 ATGCTTCAGATGCTGTTTCCCTGCAAGACATCGAGAGATGAATTCAGATTTCACATT 2014  
 Db 1618 ATGAATCTGATGCTTTTCCCTGCGGCAATGCGAGAGATGAATTCAGATTGACATT 1677  
 QY 2015 TGTTCACGCACGGAGCAGAGACTGTTCAAGTGTCTCCCG--GCGTGCAGATCT 2071  
 Db 1678 TGTTCACGCACGGAGTCAATGAGTGTCTCCCGTGTCAAGATCTCAACCC 1737  
 QY 2072 CAACCGTCTGTCAGAAAGAGACGTATCGGAAACTCTGTGCCATTATCATCTGCTGGG 2131  
 Db 1738 GTGTCTGTCGTACAGAAAGCGAGCTATCAGAAACTGTGTCCGATTATCATCATGAGG 1797  
 QY 2132 CGGCTCTCGAGATTGCTTGTGCTCGGCTGCGATCTGTCAACGTGACCTGATGACTGT 2191  
 Db 1798 AGGCGCGCCGAGGTGGCTGCTGCGGCTGCGAAGTGGCAATGTGGAATTGATGACTGT 1857  
 QY 2192 GTTCTGAGCAATTA 2206  
 Db 1858 GACATGAAACAATTA 1872

RESULT 10

US-09-532-594B-15  
 ; Sequence 15, Application US/09532594B  
 ; Patent No. 6468524  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Chordin, John A.  
 ; APPLICANT: Kotlin, Robert M.  
 ; APPLICANT: Safer, Brian  
 ; APPLICANT: Davidson, Beverly  
 ; TITLE OF INVENTION: AAV4 VECTOR AND USES THEREOF  
 ; FILE REFERENCE: 14014.0252U2  
 ; CURRENT APPLICATION NUMBER: US/09/532.594B  
 ; CURRENT FILING DATE: 2000-03-22  
 ; NUMBER OF SEQ ID NOS: 22  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 15  
 ; LENGTH: 1872  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence; No. 6468524e =  
 ; NAME/KEY: misc feature  
 ; OTHER INFORMATION: AAV4 Rep 78 gene  
 US-09-532-594B-15

Query Match 30.6%; Score 1444.8; DB 4; Length 1872;  
 Best Local Similarity 86.4%; Pred. No. 0;  
 Matches 1620; Conservative 0; Mismatches 249; Indels 6; Gaps 2;  
 QY 335 ATGCCGGGCTTCTACGAGATCGTGAATCAAGGTGCCGAGCGACTGGAACGACCTGCCG 394  
 Db 1 ATGCCGGGCTTCTACGAGATCGTGAAGGTGCCGAGCGACTGGAACGACCTGCCG 60  
 QY 395 GGCATTCTGACTCGTTGTGAGCTGGGTGGCCGAGAAAGGATGGAGCTGCCCGCGAT 454  
 Db 61 GGCATTCTGACTCTTTGTGAGCTGGGTGGCCGAGAAAGGATGGAGCTGCCCGCGAT 120  
 QY 455 TCTGACATGATCTGAATCTGATTGAGAGGACCCCTGACCGTGGCCGAGAGCTGACG 514  
 Db 121 TCTGACATGATCTGAATCTGATTGAGAGGACCCCTGACCGTGGCCGAGAAAGCTGCAA 180  
 QY 515 CGCGACTTCTGTGTCATGAGCGCGCGGTGAGTAAGGCCCGAGGACCTCTTCTTTGTT 574  
 Db 181 CGCGAGTCTGTGTCAGAGTGGCGCGGTGAGTAAGGCCCGAGGACCTCTTCTTTGTC 240  
 QY 575 CAGTTGAGAAAGGCGAGTCTTACTTCCACTTCATATTTCTGTGAGAGACCAAGGCGTC 634

Db 241 CAGTTGAGAAAGGCGAGTCTTACTTCCACTGCACTCTGTGAGAGACCGTGGCGTC 300  
 QY 635 AAATTCATGCTGCTGGCGCTTCTGAGTCAATTAAGGACCAAGCTGTGACAGACATC 694  
 Db 301 AAATTCATGCTGCTGGCGCTTCTGAGTCAATTAAGGACCAAGCTGTGACAGACATC 360  
 QY 695 TACCGCGGATCGAGCCGACCTGCCCCAATGTTCCGCGTGAACCAAGCGCTAATGTC 754  
 Db 361 TACCGCGGATCGAGCCGACCTGCCCCAATGTTCCGCGTGAACCAAGCGCTAATGTC 420  
 QY 755 GCCGAGGCGGGAACAAGGTGTGAGAGAGTGTACATCCCACTACCTCTGCGCAAG 814  
 Db 421 GCCGAGGCGGGAACAAGGTGTGAGAGAGTGTACATCCCACTACCTCTGCGCAAG 480  
 QY 815 ACTGACCCCGAGCTGCACTGGCGGTGAGTACATGAGAGATATATAAGCGCTGTTG 874  
 Db 481 ACCGACCCCGAGCTGCACTGGCGGTGAGTACATGAGAGATATATAAGCGCTGTTG 540  
 QY 875 AACCTGGCGGAGCGCAACCGGTGTGCGGAGCAGCAGCAGTCAAGCAGACCGAG 934  
 Db 541 AATCTCGGAGCGTAACCGGTGTGCGGAGCAGCAGCAGTCAAGCAGCAGCAG 600  
 QY 935 GAGCAGAACAGGAAATCTGAACCCCAATCTGACGCGCTGTATCCGTCAAAAAC 994  
 Db 601 GAGCAGAACAGGAAATCTGAACCCCAATCTGACGCGCTGTATCCGTCAAAAAC 660  
 QY 995 TCCGCGGCTTACATGAGTGTGTGCGGTGCTGTGAGACCGGCGATCACCTCCGAGAG 1054  
 Db 661 TCCGCGGCTTACATGAGTGTGTGCGGTGCTGTGAGACCGGCGATCACCTCCGAGAG 720  
 QY 1055 CAGTGGATCCAGAGAGACCAAGGCTCTGATCATCTCTTCAACGCGCTTCCACTCGCG 1114  
 Db 721 CAATGGATCCAGAGAGACCAAGGCTCTGATCATCTCTTCAACGCGCTTCCACTCGCG 780  
 QY 1115 TCCGATCAAGGCGCTGTGCAATGCGGCAAGATGAGCGCTGACCAATCCGCG 1174  
 Db 781 TCACAAATCAAGGCGCTGTGCAATGCGGCAAGATGAGCGCTGACCAATCCGCG 840  
 QY 1175 CCCGACTACCTGTAGGCGCGCTGCGCGCGGCGGCAATTAACCAACCGCATCTACCG 1234  
 Db 841 CCCGACTACCTGTAGGCGCGCTGCGCGCGGCGGCAATTAACCAACCGCATCTACCG 900  
 QY 1235 ATCTGAGTGAACGCGCTGCAACCTGCTACGCGCGCTCTTCTCGGCTGGCG 1294  
 Db 901 ATCTGAGTGAACGCGCTGCAACCTGCTACGCGCGCTCTTCTCGGCTGGCG 960  
 QY 1295 CAGAAAGCTTCGGAAGCGCAACCATCTGCTGTTGGCGCGGCGGCAAG 1354  
 Db 961 CAGAAAGCTTCGGAAGCGCAACCATCTGCTGTTGGCGCGGCGGCAAG 1020  
 QY 1355 ACCAATTCGCGGAAGCGCATCGCCACGCGCTGCTTCAAGGCTGCTCACTGAC 1414  
 Db 1021 ACCAATTCGCGGAAGCGCATCGCCACGCGCTGCTTCAAGGCTGCTCACTGAC 1080  
 QY 1415 AATGAGAACTTCTCTTCAATGATGCTGCAAGATGATGTTGAGAGAGGCG 1474  
 Db 1081 AATGAGAACTTCTCTTCAATGATGCTGCAAGATGATGTTGAGAGAGGCG 1140  
 QY 1475 AAGATGACGCGCAAGGTGCTGAGTCCGCAAGGCGCATCTCGCGGCGAGCAAGTGGC 1534  
 Db 1141 AAGATGACGCGCAAGGTGCTGAGTCCGCAAGGCGCATCTCGCGGCGAGCAAGTGGC 1200  
 QY 1535 GTGACCAAAAGTGCAAGTGTCCGCGGAGATGACCCCAACCCCGTATGCTACCTCC 1594  
 Db 1201 GTGACCAAAAGTGCAAGTGTCCGCGGAGATGACCCCAACCCCGTATGCTACCTCC 1260  
 QY 1595 AACCAACATATGCGCGCTGATGACGGAACAGACCACTTGAAGCAGCAGAGCG 1654  
 Db 1261 AACCAACATATGCGCGCTGATGACGGAACAGACCACTTGAAGCAGCAGAGCG 1320  
 QY 1655 TTGAGGACCGGATGTTCAATTGAATCACTCACCGCGCTGTGAGCATGCTTGGCAAG 1714

Db 1321 CTCAGGACCGGATGTTCAAGTTGAGCTCACCAGCGCCTGGAGCAGCACTTGGCAG 1380  
QY 1715 GTGACAAAGCAGAGTCAAAAGAGTTCTCCGCTGGGCGCAGATCAGTGACCGAGGTG 1774  
Db 1381 GTCACCAAGCAGAGTCAAAAGAGTTCTCCGCTGGGCGCAGATCAGTGACCGAGGTG 1440  
QY 1775 GCGCATGAGTTCTACGTCAAGAAAGGTGAGCAACAAAGACCCGCGCCGATGACGCG 1834  
Db 1441 ACTACAGAGTTTACGTCAAGAAAGGTGAGCTAGAAAGAGCGCCGCGCCCAATGACGCA 1500  
QY 1835 GATAAAGCAGCGCCCAAGCGGCGCTGCCCTCAGTCGCGGATCCATGACGTACAGCGC 1894  
Db 1501 GATATAAGTAGAGCCCAAGCGGCGCTGCTCCGTCACTTGGCAGCCATCGACGTACAGCGC 1560  
QY 1895 GAAGAGCTCCGCTGGAATTGCGCAGACAGTACCAAAACAAATGTTCTGTCACGCGGCG 1954  
Db 1561 GA--AGCTCCGCTGGAATAAGCGGACAGGTACCAAAACAAATGTTCTGTCACGCGGT 1617  
QY 1955 ATGCTTCAAGTCTGTTTCCCTGCAAGACATGCGAGAGATGATCAGATTTCACATT 2014  
Db 1618 ATGAATCTGATGTTTCCCTGCGCAATGCGAGAGATGATCAGATTTCAGATT 1677  
QY 2015 TGCTTACGACGCGGACGAGAGACTGTTCAAGTCTTCCCG--GCGTGTCAAGTCT 2071  
Db 1678 TGCTTACGACGCGGCTCATGAGACTGTGCGAGTCTTCCCGTGTCAAGTCTCAACCC 1737  
QY 2072 CAACCGTCTGCAAGAAAGAGACGTATCGGAACTCTGTGCCATTCAATCTGTGCGG 2131  
Db 1738 GTGCTGTGCTGCAAGAAAGCGAGCATGATCAAGAACTGTCCGATTCAATCATCATGCGG 1797  
QY 2132 CGGCTCCCGAGATTGCTGCTGCGCTGCGCTGCGTCAACGTGACCTGGATGACTGT 2191  
Db 1798 AGGCGCGCCGAGGTGCGCTGCTGCGCTGCGCACTGCGCAATGTGACTTGGATGACTGT 1857  
QY 2192 GTTCTGAGCAATTA 2206  
Db 1858 GACATGGAACATTA 1872

## RESULT 11

US-08-331-384-2/c

; Sequence 2, Application US/08331384

; Patent No. 5856152

; GENERAL INFORMATION:

; APPLICANT: Wilson, James M.

; APPLICANT: Kelley, William M.

; APPLICANT: Fisher, Krishna J.

; TITLE OF INVENTION: Hybrid Adenovirus-AAV Vector and Methods

; TITLE OF INVENTION: of Use Thereof

; NUMBER OF SEQUENCES: 2

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Howson and Howson

; STREET: Spring House Corporate Cntr, P.O. Box 457

; CITY: Spring House

; STATE: Pennsylvania

; COUNTRY: USA

; ZIP: 19477

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/331,384

; FILING DATE:

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Bak, Mary E.

; REGISTRATION NUMBER: 31,215

; REFERENCE/DOCKET NUMBER: UPNG1149USA

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 215-540-9200

; TELEFAX: 215-540-5818

; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4910 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: unknown  
; MOLECULE TYPE: cDNA  
US-08-331-384-2

Query Match 30.2%; Score 1425.8; DB 2; Length 4910;  
Best Local Similarity 83.8%; Pred. No. 0;  
Matches 1687; Conservative 0; Mismatches 312; Indels 14; Gaps 6;

QY 205 GTCTGTATTAGCTGTACAGTGAAGTCTTTGCGACATTTTGACACACCGTGGCATT 264  
Db 2646 GTCTGTATTAGAGTCAAGTGAAGTCTTTGCGACATTTTGACACACCGTGGCATT 2588  
QY 265 TAGGTAATATAGCGGAGTGAAGTGAAGTCTTCCATTG-ACCGGAATTGAAAC 323  
Db 2587 CTGGTAATTAAGCGCGAGTGAAGTGAAGTCTTCCATTGAAAGCGGAGTTGAAAC 2528  
QY 324 GAGCAGACCATATGCGCGGCTTACAGATCTGATCAAGTCCGAGCGACCTGAGC 383  
Db 2527 GCGCAGCGCCATGCGCGGCTTACAGATCTGATCAAGTCCGAGCGACCTGAGC 2468  
QY 384 AGCAGTCCGCGGCTTCTGACTGTTTGTGAGTGGTGGCGGAGGAATGGAGC 443  
Db 2467 GGCATCTGCGCGGCTTCTGACTGTTTGTGAGTGGTGGCGGAGGAATGGAGT 2408  
QY 444 TGCCCCCGGATTTGACATGATCTGAATCTGATTGAGCAGGACCCCTGACCGTGCGC 503  
Db 2407 TGCGCGAGATTTGACATGATCTGAATCTGATTGAGCAGGACCCCTGACCGTGCGC 2348  
QY 504 AGAAGTCAAGCGGCTTCTGACTGTTTGTGAGTGGTGGCGGAGGAATGGAGC 563  
Db 2347 AGAAGTCAAGCGGCTTCTGACTGTTTGTGAGTGGTGGCGGAGGAATGGAGC 2288  
QY 564 TCTTCTTGTTCAGTTGAGAAAGGCGAGTCTTCTTCCACTTCCATATTCTGTGAGAG 623  
Db 2287 TTTTCTTGTTCAGTTGAGAAAGGCGAGTCTTCTTCCACTTCCATATTCTGTGAGAG 2228  
QY 624 CCACGGGGTCAATATCCATGCTGCGCGCGCTTCTTCTGAGTCAATTAAGGACAGCTGG 683  
Db 2227 CCACGGGGTCAATATCCATGCTGCGCGCGCTTCTTCTGAGTCAATTAAGGACAGCTGG 2168  
QY 684 TGCAGACATCTACCGCGGATCGAGCGGACCCCTGCGCACTGCTGCGGAGCAAGA 743  
Db 2167 TTCAGAGATTACCGCGGATCGAGCGGACCCCTGCGCACTGCTGCGGAGCAAGA 2108  
QY 744 CGCTAATGCGCGCGGAGGCGGAGCAAGGTGTGAGCAGTGTACATCCCACTACC 803  
Db 2107 CCAGAAATGCGCGCGGAGGCGGAGCAAGGTGTGAGTGTACATCCCACTACC 2048  
QY 804 TCTGCGCAAGTCAAGCGGATCGAGTGGCGGTGAGTCAATGAGAGATATATAA 863  
Db 2047 TGCTCCCAAAACCCAGCTGAGTCCAGTGGCGGTGAGTCAATGAGAGATATATAA 1988  
QY 864 GCGCTGTTTGAACCTGCGCGGAGGCGGAGGCTGCGCGGAGGAGCTGACCGACGTCA 923  
Db 1987 GCGCTGTTTGAATCTACGAGGAGGCGGAGGCTGCGCGGAGGAGCTGACCGACGTGT 1928  
QY 924 GCGCAGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 983  
Db 1927 GCGCAGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1868  
QY 984 GGTCAAAAACCTCGCGGCTTACATGAGCTGTGCGGCTGCTGCTGCTGCTGCTGCTG 1043  
Db 1867 GATCAAAAACCTTACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1808  
QY 1044 CCTCCGAGAGCAGTGAATCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1103  
Db 1807 CCTCGAGAGAGCAGTGAATCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1748

QY 1104 CCAATCGGGTCCAGATCAAGCCGCTCTGGACAATGCCGGCAGATCATGCGCTGA 1163  
DB 1747 CCAATCGGGTCCCAATCAAGCTGCTTGGACAATGCCGGAAGATTATGAGCTGA 1688  
QY 1164 CCAATCGGGTCCAGATCAAGCCGCTCTGGACAATGCCGGCAGATCATGCGCTGA 1223  
DB 1687 CTAAGACCGCCCGGACTACTGCTGGGCGCAGACGCCGCTGAGAGACATTTCCAGCAATC 1628  
QY 1224 GCATCTACCGCATCTGAGCTGAAGCGCTACGAACCTGCTACGCGGCTCCGCTTTC 1283  
DB 1627 GGATTATATAATTTTGAAGTAAACGGGTACGATCCCAATATGCGGCTTCCGCTTTC 1568  
QY 1284 TCGGCTGGGCGGCAAGAAAGTTCCGGAAGCGCAACACCATCTGGCTTTGGGCGGCGCA 1343  
DB 1567 TGGGATGGGCGCAGAAAAAGTTCCGCAAGAGGAAACACCATCTGGCTTTGGGCGCTGCA 1508  
QY 1344 CCACGGGCAAGCAACATCGCGGAAGCCATCGCCACGCGGCTTCTACGCGCTGCG 1403  
DB 1507 CTACCGGGAAGCAACATCGCGGAGCCATAGCCACACATGTCCTTCTACGCGCTGCG 1448  
QY 1404 TCAACTGGAACCAATGAGAACTTCCCTTCAATGATTGCGTGAAGAGATGATCTGT 1463  
DB 1447 TAACTGGAACCAATGAGAACTTCCCTTCAACGACTGTGTGAGCAAGATGATCTGT 1388  
QY 1464 GGGAGAGGGCAAGATGACGCGCAAGCTGGAAGTCCGCCAAGCCATCTCGCGGCA 1523  
DB 1387 GGGAGAGGGCAAGATGACGCGCAAGCTGGAAGTCCGCCAAGCCATCTCGGAGGA 1328  
QY 1524 GCAAGTGGCGGCTGGAGCCAAAAGTCAAGTCTCGGCGGCAAGATGACGCGGCTGA 1583  
DB 1327 GCAAGTGGCGGCTGGAGCCAAAAGTCAAGTCTCGGCGGCAAGATGACGCGGCTGA 1268  
QY 1584 TCGTCACTCCCAACCAACATGTCGCGGCTGATTGAGCGGGAAGACCAACCTTGAGC 1643  
DB 1267 TCGTCACTCCCAACCAACATGTCGCGGCTGATTGAGCGGGAAGACCAACCTTGAGC 1208  
QY 1644 ACCAGAGCGGCTGAGGAGCCGATGTTCAAAATTGAATCAACGCGCTGAGCAG 1703  
DB 1207 ACCAGAGCGGCTGAGGAGCCGATGTTCAAAATTGAATCAACGCGCTGAGCAG 1148  
QY 1704 ACTTGGCAAGGTGACAAAGCAGAAAGTCAAAAGTCTTCCGCTGGCGCAGATCAG 1763  
DB 1147 ACTTGGCAAGGTGACAAAGCAGAAAGTCAAAAGTCTTCCGCTGGCGCAGATCAG 1088  
QY 1764 TGACCGAGTGGCGCATGAGTCTTACGTCAGAAAGGCTGAGGCCAACAAGACCGGCGC 1823  
DB 1087 TGCTTGAAGTGAAGCATGAACTTCTACGTCAGAAAGGCTGAGGCCAACAAGACCGGCGC 1028  
QY 1824 CCGATGACCGGATTAAGAGCGAGCCCAAGCGGCTGCGCTCAGTCGCGGATCCATGA 1883  
DB 1027 CCGATGACCGGATTAAGAGCGAGCCCAAGCGGCTGCGCTCAGTCGCGGATCCATGA 968  
QY 1884 CGTCAAGCGGAAAGAGCTCCGCTGAGCTTTCGCGAGAGGTACCAAAACAATGTTCTC 1943  
DB 967 CGTCAAGCGGAAAGAGCTCCGCTGAGCTTTCGCGAGAGGTACCAAAACAATGTTCTC 911  
QY 1944 GTACGCGGCGATGCTTCAAGATGCTTTCCTGCAAGACATGCGAGAGATGAATCAGA 2003  
DB 910 GTACGCGGCGATGCTTCAAGATGCTTTCCTGCAAGACATGCGAGAGATGAATCAGA 851  
QY 2004 ATTCAACATTTGCTTCAAGACGCGAGAGAGACTGTTCAAGATGCTTCCCGCGCTGT 2063  
DB 850 ATTCAACATTTGCTTCAAGACGCGAGAGAGACTGTTCAAGATGCTTCCCGCGCTGT 794  
QY 2064 CAGATCTCAACCGGT--CGTCAAGAAAGAGAGCTATCGGAACCTGTGCGATTCATC 2120  
DB 793 CAGATCTCAACCGGT--CGTCAAGAAAGAGAGCTATCGGAACCTGTGCGATTCATC 734  
QY 2121 ATCTGCTGGGCGGCTCCGAGATGCTTGTGCGCTGCGATGCTCAACGTGAGC 2180  
DB 733 ATATCATGGGAAGGTGCCAGA--CGCTGCACTGCGCTGCGATGCTCAATGTGAT 677  
QY 2181 TGATGACTGTGTTCTGAGCAATAAATGACTT 2213

DB 676 TGATGACTGCATCTTTGACAAATAATGATTT 644

RESULT 12  
US-08-836-087-2/c  
Sequence 2, Application US/08836087  
Patent No. 5871982  
GENERAL INFORMATION:  
APPLICANT: Trustees of University of Pennsylvania  
APPLICANT: Wilson, James M.  
APPLICANT: Kelley, William M.  
APPLICANT: Fisher, Krishna J.  
TITLE OF INVENTION: Hybrid Adenovirus-AAV Vector and  
TITLE OF INVENTION: Methods of Use Thereof  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Howson and Howson  
STREET: Spring House Corporate Cntr, PO Box 457  
CITY: Spring House  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19477  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/836,087  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/331,384  
FILING DATE: 28-OCT-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Bak, Mary E.  
REGISTRATION NUMBER: 31,215  
REFERENCE/DOCKET NUMBER: GNVN.007PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-540-9200  
TELEFAX: 215-540-5818  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4910 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: cDNA  
US-08-836-087-2

Query Match 30.2%; Score 1425.8; DB 2; Length 4910;  
Best local Similarity 83.8%; Pred. No. 0;  
Matches 1687; Conservative 0; Mismatches 312; Indels 14; Gaps 6;

QY 205 GTCTGTATTAGCTGTACGCTGAGTGTCTTTGGCAATTTTGGACACCAACGCTGCCATT 264  
DB 2646 GTCTGTATTAGAGGTACGCTGAGTG-TTTGGCAATTTTGGACACCAACGCTGCCATT 2588  
QY 265 TAGGATATATGCGCGAGTGAAGCAGCAGGATCTCCATTTG-ACCGGAATTTGAAC 323  
DB 2587 CTGGGTATTAAAGCCGAGTGAAGCAGCAGGCTCTCCATTTTGAAGCGGAGTTGAAC 2528  
QY 324 GAGCAGACCCATGCGGCTTCTACGAGATCGTATGATGAGTCCGAGCAGCCTGACG 383  
DB 2527 GCGCAGCCCGCATGCGGCTTCTACGAGATGTGATTAAGTCCCGAGCAGCCTGACG 2468  
QY 384 AGCACTGCGCGCATTTCTGACTGTTGTGAGCTGGGTGCGCGAGAGAAATGGAGC 443  
DB 2467 GGCATCTGCCCGCATTTCTGACAGCTTTGTGAAGTGGGTGCGCGAGAGATGGAGT 2408  
QY 444 TGCCCCCGGATTTGACATGATCTGAATCTGATGAGCAGCAGCAGCAGCAGCAGCAGCAG 503



Db 2407 TGCCGCCAGATTCTGACATGATCTGAATCTGATTGAGCAGGACCCCTGACCCGTGGCCG 2348  
QY 504 AGAAGCTGACGCGGCACTTCTGTGTCCAATGGCCCGCTGAGTAAGGCCCGGAGGCC 563  
Db 2347 AGAAGCTGACGCGGCACTTCTGTGACGGAATGGCCCGTGTGAGTAAGGCCCGGAGGCC 2288  
QY 564 TCTTCTTTGTTGACGTTGAGAGGGCGAGTCTTAATCTCCATCTTCTGTGAGAGA 623  
Db 2287 TTTTCTTTGTGCAATTTGAGAGGAGAGAGCTACTTCCACATGACAGTGTCTGTGAAA 2228  
QY 624 CCACGGGGGTCAAAATCCATGTGTGTGGCCGCTTCTGAGTCAAGTATAGGACAAGCTGG 683  
Db 2227 CCACGGGGGTCAAAATCCATGTGTGTGGGACGTTTCTGAGTCAAGTATAGGACAAGCTGG 2168  
QY 684 TGCAGACCATCTACCGCGGGGATCGAGCCGACCTGCCCCAATGTGTGCGGCTGACCAAGA 743  
Db 2167 TTTAGAGAAATTTACCGCGGGATCGAGCCGACTTTGCCAAACTGGTTCGCGGTCACAAGA 2108  
QY 744 CGCGTAATGGCGCGGAGGGGGGAAACAAGGTGTGTGAGAGAGTGTACATCCCAACTACC 803  
Db 2107 CCAGAAATGGCGCGGAGCGGGAACAAGGTGTGTGAGTGTGCTACATCCCAATTACT 2048  
QY 804 TCCTGCCCAAGACTCAGCCCGAGCTGAGTGGCGGTGAGTAAACATGAGAGATATATA 863  
Db 2047 TGCTCCCAAAACCAACCTGAGCTCCAGTGGCGGTGAGTAAATATGAAACAGTATTTAA 1988  
QY 864 GCGCTGTGTTGAATCTGCGCGGAGCGCAACCGCTGTGGCGCAGCACTGACCACTCA 923  
Db 1987 GCGCTGTGTTGAATCTCAGGAGCGTAAACGTTGTGGCGCAGCATCTGACGACAGTGT 1928  
QY 924 GCCAGACCCAGAGAGAGAAACAAGGAATCTGAACCCCAATTTGACCGCGCTGTCTATCC 983  
Db 1927 CGCAGACCGCAGAGAGAGAAACAAGGAATCTGAATCTGATGCGCGGTGATCA 1868  
QY 984 GGTCAAAAACCTCCGCGCGCTACATGAGCTGTGGGTGGTGTGTGACCGGGGATCA 1043  
Db 1867 GATCAAAAACCTTCAAGCTACGATGAGCTGTGGGTGGTGTGTGACAGGGGATTA 1808  
QY 1044 CTTCCGAGAGAGAGTGTATCCAGAGAGACAGGCTGTATCATCTCTTCAACGCGCTT 1103  
Db 1807 CTTGAGAGAGAGTGTATCCAGAGAGACAGGCTGTATCATCTCTTCAATGCGGCTT 1748  
QY 1104 CCAACTCGCGGTCCAGATCAAGGCGCTGTGACAATGCGCGCAAGATCAATGCGCTGA 1163  
Db 1747 CCAACTCGCGGTCCCAATCAAGGCTGTGACAATGCGCGCAAGATCAATGCGCTGA 1688  
QY 1164 CCAATCCGCGCGCTACTCTGTGAGCGCCGCTCCGCGCGCGGACATTAACAACC 1223  
Db 1687 CTAAAACCGCGCGCTACTCTGTGAGCGCGCGCGCGCGCGCGGACATTTCCAGCAATC 1628  
QY 1224 GCATCTACCGCATCTGTGAGCTGAACGGCTTACGAACCTGCTACGCGCGCTCTTTC 1283  
Db 1627 GGATTTATAAATTTTGAACCTAAACGGGTACATCTCCCAATATGCGGCTTCTTTC 1568  
QY 1284 TCGGCTGGGCGCGAGAAAGGTTGCGGAAGCGCAACCATCTGCTGTTGGGCGCGCA 1343  
Db 1567 TGGATGGGCGCGAGAAAGTTGCGGAAGAGAACACCATCTGCTGTTGGGCGCTGCA 1508  
QY 1344 CCACGGGCAAGACCAATCGCGGAAGCCATCGCCACGCGCTGCTTCAAGGCTGCG 1403  
Db 1507 CTACCGGAGAACCAATCGCGGAGCGCATGCGCCACACTGTGCTTCAAGGCTGCG 1448  
QY 1404 TCAACTGACCAATGAGAACTTCTTCAATGATTTGCTGACAAAGATGATCTGT 1463  
Db 1447 TAACTGACCAATGAGAACTTCTTCAAGCACTGTGTGACAAAGATGATCTGT 1388  
QY 1464 GGGAGAGGGCAAGATGACGGCCAAAGTGTGTGAGTCCGCCAAGGCCATTTCTGGCGCA 1523  
Db 1387 GGGAGAGGGGAAAGATGACCGCCAAAGTGTGTGAGTCCGCCAAGGCCATTTCTGGAGGAA 1328  
QY 1524 GCAAGGTGCGGTGAGCAAAAAGTGTGAGTGTGCGCCAGATGAGATGAGATGAGATGAG 1583  
Db 1327 GCAAGGTGCGGTGAGCAAAAAGTGTGAGTGTGCGCCAGATGAGATGAGATGAGATGAG 1268

QY 1584 TCGTACCTCCAAACACCAACATGTGCGCCGTGATTGACGGGAAACAGACCACTTGCAGC 1643  
Db 1267 TCGTACCTCCAAACACCAACATGTGCGCCGTGATTGACGGGAACTCAACGACCTTGCAGC 1208  
QY 1644 ACCAGCAGCGGTGACAGACCGGATGTTCAATTTGAATCTACCCCGCGTGTGAGCATG 1703  
Db 1207 ACCAGCAGCGGTGACAGACCGGATGTTCAATTTGAATCTACCCCGCGTGTGAGCATG 1148  
QY 1704 ACTTGGGAAGGTGACCAAGCAGGAAGTCAAAAGATTTCTTCCGCTGGCGCAGATCACG 1763  
Db 1147 ACTTGGGAAGGTGACCAAGCAGGAAGTCAAAAGATTTCTTCCGCTGGCGCAGATCACG 1088  
QY 1764 TGACGAGGTGCGCATGAGTCTACGTCAAGAAAGGTGAGGCCAACAAAGACCCGCC 1823  
Db 1087 TGGTGTGAGGTGAGCATGAAATTTACGTCAAAAAGGTGAGGCCAAAGAAAGACCCGCC 1028  
QY 1824 CCGATGACGGGATTAAGGCGAGCCCAAGCGGCGCTGCGCTCACTGCGGATCCATCGA 1883  
Db 1027 CCAGTACGAGATATAAGTACGCCCAACGCGTGGCGGAGTGTGCGCAGCATCGA 968  
QY 1884 CGTACAGCGCGGAAGAGCTCCGCTGCACTTTGCCGACAGATACCAAAACAATGTTCTC 1943  
Db 967 CGTACAGCGCGGA--AGCTTCACTCACTACGACAGACAGATACCAAAACAATGTTCTC 911  
QY 1944 GTACGCGGCGCATGCTTCAAGTGTGTTCCCTGCAAGACATGCGAGAGATGAATCAGA 2003  
Db 910 GTACGCGGCGCATGAACTGTATGTTTCCCTGACAGACATGCGAGAGATGAATCAGA 851  
QY 2004 ATTCAACATTTGCTTCAAGCAGCGGAGAGAGACTGTTCAAGTGTGTTCCCGCGGTGT 2063  
Db 850 ATTCAACATTTGCTTCACTACGAGAGAGAGACTGTTCAAGTGTGTTCCCGCGGTGT 794  
QY 2064 CAGAAATCTCAACCGGT--CGTCAAGAAAGAGAGATGCGAAACTCTGTGCAATTCATC 2120  
Db 793 CAGAAATCTCAACCGGTCTTCTGTGCTCAAAAAGCGTATCAGAACTGTGCTACATTCATC 734  
QY 2121 ATCTGCGGCGCGGCTCCCGAGATGCTTCTGCGCTGCGATCTGTCAAGTGTGACC 2180  
Db 733 ATATCATGGGAAAGGTGCCAGA---CGTTGCACTGCTGCGATCTGTCAATGTGAT 677  
QY 2181 TGGATGACTGTGTTTCTGAGCAATTAATGACTT 2213  
Db 676 TGGATGACTGATCTTTGAACAATTAATGATTT 644

RESULT 13  
US-09-246-320-2/c  
; Sequence 2, Application US/09246320  
; Patent No. 6251677  
; GENERAL INFORMATION:  
; APPLICANT: Trustees of University of Pennsylvania  
; APPLICANT: Wilson, James M.  
; APPLICANT: Kelley, William M.  
; APPLICANT: Fisher, Krishna J.  
; TITLE OF INVENTION: Hybrid Adenovirus-AAV Vector and  
; TITLE OF INVENTION: Methods of Use Thereof  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Howson and Howson  
; STREET: Spring House Corporate Cntr, PO Box 457  
; CITY: Spring House  
; STATE: Pennsylvania  
; COUNTRY: USA  
; ZIP: 19477  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/246,320  
; FILING DATE:



Db 850 ATTCAATATCTGCTTCACTCAGCAGAGAAAGACTGTTTAGTGTCTCC--CGTGT 794

Qy 2064 CAGATCTCAACCGGT--CGTCAGAAAGAGACGTATCGGAACCTCTGTCATTCATC 2120

Db 793 CAGATCTCAACCGGTCTTCTGTCGTCAAAAAGCGGTATCAGAACTGTGCTACATTCATC 734

Qy 2121 ATCTGCTGGGCGGCTCCCGAGATTGCTTCTGCGCTCGCATCTGTCAACGTGACC 2180

Db 733 ATATCATGGGAAAGGTGCCAGA--CGCTTGACTGCTCGCATCTGTCAATGTGATT 677

Qy 2181 TGGATGACTGTGTTTCTGAGCAATAATGACTT 2213

Db 676 TGGATGACTGATCTTTGAACAATAATGATT 644

RESULT 14

US-09-546-738-2/c

Sequence 2, Application US/09546738

Patent No. 6387368

GENERAL INFORMATION:

APPLICANT: Trustees of University of Pennsylvania

Wilson, James M.

Kelley, William M.

Fisher, Krishna J.

TITLE OF INVENTION: Hybrid Adenovirus-AAV Vector and

Methods of Use Thereof

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: Howson and Howson

STREET: Spring House Corporate Cntr, PO Box 457

CITY: Spring House

STATE: Pennsylvania

COUNTRY: USA

ZIP: 19477

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/546,738

FILING DATE: 11-Apr-2000

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/246,320

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Bak, Mary E.

REGISTRATION NUMBER: 31,215

TELECOMMUNICATION INFORMATION:

TELEPHONE: 215-540-9200

TELEFAX: 215-540-5818

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 4910 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: unknown

MOLECULE TYPE: cDNA

SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-09-546-738-2

Query Match 30.2%; Score 1425.8; DB 4; Length 4910;

Best Local Similarity 83.8%; Pred. No. 0;

Matches 1687; Conservative 0; Mismatches 312; Indels 14; Gaps 6;

Qy 205 GTCCTGTATAGCTGTACGTGAGTGTCTTTGCGACATTTTGCAGACCAAGTGGCCATT 264

Db 2646 GTCCTGTATAGAGTACGTGAGTG-TTTTGCACATTTTGCAGACCAATGTGTGACG 2588

Qy 265 TAGGTATATATGCGCGAGTGAGCGAGAGATCTCCATTTTG-ACCGCAATTTGAAC 323

Db 1507 CTACCGGGAAGACCAACATCGCGAGGCCATAGCCCACTGTGCCCTTCTACGGGTGCG 1448

Db 2587 CTGGTAATTAAAGCCGAGTGAGCAGCAGGGTCTCCATTTTGAAGCGGAGTTGAAC 2528

Qy 324 GAGCAGCAGCCATGCGCGGCTTCTACAGATCGTGATCAAGGTGCCAGCAGCTGAGC 383

Db 2527 GCGCAGCGCCATGCGGGGTTTACAGATGTGATTAAGTCCCGCAGCAGCTTGAAC 2468

Qy 384 AGCAGCTGCGCGCATTTCTGATCTGTTGTGAGCTGGGTGCCGAGAAAGATGGAGC 443

Db 2467 GGCATCTGCGCGCATTTCTGACAGCTTTGTGAAGTGGGTGCCGAGAAAGATGGAGT 2408

Qy 444 TGCCCCGGGATTTCTGACATGATCTGAATCTGATTGAGCAGGACCCCTGACCGTGCCG 503

Db 2407 TGCCGAGATTTCTGACATGATCTGAATCTGATTGAGCAGGACCCCTGACCGTGCCG 2348

Qy 504 AGAGCTGACGCGGACTTCTGTCATGCGCGCGGCTTCTGAGTGAAGGCGCGGAGGCC 563

Db 2347 AGAGCTGACGCGGACTTCTGTCATGCGCGCGGCTTCTGAGTGAAGGCGCGGAGGCC 2288

Qy 564 TCTTCTTTGTTCAAGTTCGAGAGGCGAGTCTTCTTCTGATGATTAAGGACAAGCTGG 623

Db 2287 TTTTCTTTGTTCAAGTTCGAGAGGCGAGTCTTCTTCTGATGATTAAGGACAAGCTGG 2228

Qy 624 CCAGCGGCTCAATCCATGTTGCTGGCGGCTTCTGATGATTAAGGACAAGCTGG 683

Db 2227 CCAGCGGCTCAATCCATGTTGCTGGCGGCTTCTGATGATTAAGGACAAGCTGG 2168

Qy 684 TGCAGACCATCTACCGCGGATCGAGCCGACCCCTGCGCAACTGTGCGGTGACCAAGA 743

Db 2167 TTCAGAGATTTTACCGCGGATCGAGCCGACCTTTCGCAACTGTGCGGTGACCAAGA 2108

Qy 744 CGCGTATGCGCGCGGAGGCGGGAACAAGGTGTGAGCAGAGTGTACATCCCAACTACC 803

Db 2107 CCAGAAATGCGCGCGGAGGCGGGAACAAGGTGTGAGTGTGTACATCCCAATTACT 2048

Qy 804 TCTGCGCAAGACTCAGCCCGAGCTGAGTGGCGGTGACTTAATGAGAGATATATAA 863

Db 2047 TGCTCCCAAAACCAGCTGAGCTCAAGTGGCGGTGACTTAATGAGAGATATATAA 1988

Qy 864 GCGCTGTTGAACCTGCGCGGCAAAAGGCTGCTGCGCGCAAGCTGAGCCAGCTCA 923

Db 1987 GCGCTGTTGAATCTCAGGAGGCTAAACGTTGTGTGCGCGCAAGCTGAGCCAGCTGT 1928

Qy 924 GCCAGCCAGAGCAGAGACAGAACAGAAATCTGAACCCCAATCTGACGCGCTGTATCC 983

Db 1927 CGCAGACGAGAGCAGAGACAGAACAGAAATCTGAATCCCAATCTGATGCGCGGTATCA 1868

Qy 984 GGTCAAAAACCTCGCGGCTACATGAGCTGTGCGGTGCTGTGAGACCGGGGATCA 1043

Db 1867 GATCAAAAACCTCAGCAGGATGAGAGCTGTGCGGTGCTGTGAGACCGGGGATTA 1808

Qy 1044 CCTCCGAGAGCAGTGTGATCAGAGAGACCAAGCCTCGTACATCTCTTCAACGCGCTT 1103

Db 1807 CCTCGAGAGCAGTGTGATCAGAGAGACCAAGCCTCGTACATCTCTTCAATGCGGCTT 1748

Qy 1104 CCAACTCGCGGTCCAGATCAAGCGGCTCTGACAAATGCGGCAAGATCATGCGGCTGA 1163

Db 1747 CCAACTCGCGGTCCCAATCAAGGCTGCTTGGACAAATGCGGAAAGATTATGAGCTGA 1688

Qy 1164 CCAAAATCGCGGCGGACTACTGTTAGGCGCGGCTCGCGCGGACATTAACAACCAAC 1223

Db 1687 CTAATAACGCGCGGCGGACTACTGTTAGGCGGCGGCGGCGGAGACATTTCCAGCAATC 1628

Qy 1224 GCATCTACCGCATCTGTGAGCTGAACGGCTACGAACCTGCTACGCGCGGCTCTTTC 1283

Db 1627 GGATTTATAAATTTTGAACCTAAACGGGTACGATCCCAATATGCGGCTTCCGCTTTC 1568

Qy 1284 TCGGCTGGGCGCAGAAAAGTTGCGGAAGCGCAACACCATCTGCTGTTGGGCGGCGCA 1343

Db 1567 TGGGATGGGCGCAGAAAAGTTGCGGAAGCGCAACACCATCTGCTGTTGGGCGGCGCA 1508

Qy 1344 CCACGGGCAAGCAACATCGCGGAAGCCATGCGCCACGCGGCTTCTACGGCTGCG 1403

Db 1507 CTACCGGGAAGCAACATCGCGGAGGCCATAGCCCACTGTGCCCTTCTACGGGTGCG 1448



```

QY 1404 TCACTGACCAATGAGAACTTCCCTTCAATGATTCGCTGACAGATGATCTGCT 1463
Db 1447 TAACTGACCAATGAGAACTTCCCTTCAACGACTGTGTGACAGATGATCTGCT 1388
QY 1464 GGGAGAGGGCAAGATGACGGCCAAAGTCTGTGAGTCCGCCAAGGCCATTCTGGCGCA 1523
Db 1387 GGGAGAGGGCAAGATGACGGCCAAAGTCTGTGAGTCCGCCAAGGCCATTCTGGAGAA 1328
QY 1524 GCAAGGTGCGCGTGGACCAAAAGTCAAGTCTGCGCCAGATGACCCCAAGCCCGTGA 1583
Db 1327 GCAAGGTGCGCGTGGACCAAAAGTCAAGTCTGCGCCAGATGACCCCAAGCCCGTGA 1268
QY 1584 TCGTCACTCCCAACCAACATGTGCGCGGTGATGACGGGAACAGACCACTTCGAGC 1643
Db 1267 TCGTCACTCCCAACCAACATGTGCGCGGTGATGACGGGAACAGACCACTTCGAGC 1208
QY 1644 ACCAGCAGCGGTGGAGACCGGATGTTCAAAATTGAATCAACCCCGCTGAGCATG 1703
Db 1207 ACCAGCAGCGGTGGAGACCGGATGTTCAAAATTGAATCAACCCCGCTGAGCATG 1148
QY 1704 ACTTGGCAAGTGAACAAGAGAGTCAAGAGTCTTCCGCTGGCGGAGGATCAGC 1763
Db 1147 ACTTGGCAAGTGAACAAGAGAGTCAAGAGTCTTCCGCTGGCGGAGGATCAGC 1088
QY 1764 TGACCGAGGTGCGCATGAGTCTACGTCAGAAAGGTGAGCCCAAAAGAGCCCGCC 1823
Db 1087 TGCTGAGGTGAGCATGAATTTCTACGTCAAAGAGGTGAGCCCAAAAGAGCCCGCC 1028
QY 1824 CCGATGACGCGGATAAAGCGAGCCCAAGCGGCTGCCCTCAGTCGCGGATCCATCA 1883
Db 1027 CCGATGACGCGGATAAAGCGAGCCCAAGCGGCTGCCCTCAGTCGCGGATCCATCA 968
QY 1884 CGTACAGCGCGGAGAGAGTCCGCTGAGCTTCCGACAGAGTACCAAAACAATGTTCTC 1943
Db 967 CGTACAGCGCGGAGAGTCCGCTGAGCTTCCGACAGAGTACCAAAACAATGTTCTC 911
QY 1944 GTACGCGCGGATGCTTCAAGATGCTTCCCTGCAAGACATGCGGAGAAATGAATCAGA 2003
Db 910 GTACGCGCGGATGCTTCAAGATGCTTCCCTGCAAGACATGCGGAGAAATGAATCAGA 851
QY 2004 ATTTCAACATTGCTTCAAGCAGCGGAGAGAGTGTTCAGAGTCTTCCCGCGGTGT 2063
Db 850 ATTTCAACATTGCTTCAAGCAGCGGAGAGAGTGTTCAGAGTCTTCCCGCGGTGT 794
QY 2064 CAGAACTCAACCGGT--CGTCAGAAAGAGAGTATCGAAACTGTGCGCATTCATC 2120
Db 793 CAGAACTCAACCGGTCTTCTGCTCAAAAGCGGTATCAGAAACTGTGCTACATTCATC 734
QY 2121 ATCTGCTGGGGGGCTCCCGAGATGCTGCTCGGCTGCGATCTGTCAACGTGACC 2180
Db 733 ATATCATGGGAAAGGTGCCAGA--CGCTTGACATGCTGCGATCTGTCAATGTGATT 677
QY 2181 TGATGACTGTGTTTCTGAGCAATAATGACTT 2213
Db 676 TGATGACTGTGTTTCTGAGCAATAATGATT 644

```

RESULT 15

```

US-09-532-594B-14
; Sequence 14, Application US/09532594B
; Patent No. 6468524
; GENERAL INFORMATION:
; APPLICANT: Chorini, John A.
; APPLICANT: Kotin, Robert M.
; APPLICANT: Safer, Brian
; APPLICANT: Davidson, Beverly
; TITLE OF INVENTION: AAV4 VECTOR AND USES THEREOF
; FILE REFERENCE: 14014.025202
; CURRENT APPLICATION NUMBER: US/09/532,594B
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0

```

```

; SEQ ID NO 14
; LENGTH: 1611
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence; No. 6468524e =
; NAME/KEY: misc feature
; OTHER INFORMATION: AAV4 Rep 68 gene
US-09-532-594B-14

```

```

Query Match 26.8%; Score 1266.2; DB 4; Length 1611;
Best Local Similarity 87.7%; Pred. No. 1.6e-291;
Matches 1395; Conservative 0; Mismatches 193; Indels 3; Gaps 1;

```

```

QY 335 ATGCGGGCTTCTACGAGATCGTGAAGTGTCCGAGCGGACCTGAGCAGACCACTGCCG 394
Db 1 ATGCGGGGTTCTACGAGATCGTGAAGTGTCCGAGCGGACCTGAGCAGACCACTGCCG 60
QY 395 GGCATTCTGACTCGTTGTGAGCTGGGTGGCCGAGAAAGATGGAGCTGCCCGGAT 454
Db 61 GGCATTCTGACTCTTTGTGAGCTGGGTGGCCGAGAAAGATGGAGCTGCCCGGAT 120
QY 455 TCTGACATGATCTGAATCTGATTGACAGCAGCAGCCCTGACCGTGGCGGAGAGTGCAG 514
Db 121 TCTGACATGATCTGAATCTGATTGACAGCAGCAGCCCTGACCGTGGCGGAGAGTGCAG 180
QY 515 CGGACTTCTGCTTCAATGCGCGCGGTGAGTAAGGCCCCGAGGCCCTCTTCTTGT 574
Db 181 CGGAGTCTCTGCTGAGTGGCGCGCGGTGAGTAAGGCCCCGAGGCCCTCTTCTTGT 240
QY 575 CAGTTGAGAAAGGGGAGTCTTCTTCACTTCACTTCACTTCACTTCACTTCACTTCACTT 634
Db 241 CAGTTGAGAAAGGGGAGTCTTCTTCACTTCACTTCACTTCACTTCACTTCACTTCACTT 300
QY 635 AAATCCATGCTGTGGGCGCTTCTGAGTCAAGATTAGGAGCAAGCTGTGACAGCATC 694
Db 301 AAATCCATGCTGTGGGCGCTTCTGAGTCAAGATTAGGAGCAAGCTGTGACAGCATC 360
QY 695 TACCGCGGATGAGCGGACCTGCGCACTGTTGCGCGGTGAGCAAGAGCGGTATGCG 754
Db 361 TACCGCGGATGAGCGGACCTGCGCACTGTTGCGCGGTGAGCAAGAGCGGTATGCG 420
QY 755 GCGGAGGGGAGAAAGAGTGTGAGAGTGTCTACATCCCACTACCTCTGCCCAAG 814
Db 421 GCGGAGGGGAGAAAGAGTGTGAGAGTGTCTACATCCCACTACCTCTGCCCAAG 480
QY 815 ACTGACCGGAGCTGCAAGTGGCGGTGAGTAACTGAGAGAGTATTAAGCGCTGTTG 874
Db 481 ACCGAGCGGAGCTGCAAGTGGCGGTGAGTAACTGAGAGAGTATTAAGCGCTGTTG 540
QY 875 AACCTGGCGGAGCGCAACCGGCTGTGGCGGAGCACTGAGCCAGTCAAGAGCCAG 934
Db 541 AACCTGGCGGAGCGTAAACGGCTGTGGCGGAGCACTGAGCCAGTCAAGAGCCAG 600
QY 935 GAGCAGAAAGAGAGTGAATCTGAACCCCAATTCTGACGCGCTGTATCCGGTCAAAAACC 994
Db 601 GAGCAGAAAGAGAGTGAATCTGAACCCCAATTCTGACGCGCTGTATCCGGTCAAAAACC 660
QY 995 TCCGCGCTACATGAGCTGTGCGGTGGCTGTGAGACCGGGGATCACTCCGAGAG 1054
Db 661 TCCGCGCTACATGAGCTGTGCGGTGGCTGTGAGACCGGGGATCACTCCGAGAG 720
QY 1055 CAGTGATCCAGAGAGAGGCTCTGATCATCTCTTCAACCGGCTTCCAACTCGCG 1114
Db 721 CAGTGATCCAGAGAGAGGCTCTGATCATCTCTTCAACCGGCTTCCAACTCGCG 780
QY 1115 TCCGAGATCAAGCGGCTTGTGACATGCGGCAAGATGAGCGCTGACCAATCGCG 1174
Db 781 TCCGAGATCAAGCGGCTTGTGACATGCGGCAAGATGAGCGCTGACCAATCGCG 840
QY 1175 CCGGACTACTGTAGGCGCGCTCGCGCGGAGCATTAACCAACGCGATCTACCGC 1234

```

Db 841 CCGACTACCTGGTGGGCCAGAACCCGCGGAGGACATTTCCAGCAACCGCATCTACCGA 900  
 QY 1235 ATCTGAGCTGAACGGCTACGAACCTGCTACCGGCTCCGTCTTCTCGGCTGGCC 1294  
 Db 901 ATCTCGAGATGAACGGGTACGATCCGCAATACGGGCTCCGTCTTCTCGGCTGGCC 960  
 QY 1295 CAGAAAAGTTCCGGGAAGCCCAACACCATCTGGCTTTTGGCCGGCCACACGGGCAAG 1354  
 Db 961 CAAAAGAGTTCCGGGAAGGAGAACACCATCTGGCTTTTGGCCGGCCACACGGGTAA 1020  
 QY 1355 ACCAACATCGCGGAAGCCATCGCCACGCGGCTTCTACGGCTGCGTCAACTGGACC 1414  
 Db 1021 ACCAACATCGCGGAAGCCATCGCCACGCGGCTTCTACGGCTGCGTCAACTGGACC 1080  
 QY 1415 AATGAGACTTCCCTTCAATGATTGCGTCAAGATGATGATCTGTTGGAGGAGGGC 1474  
 Db 1081 AATGAGACTTCCCTTCAACGATTGCGTCAAGATGATGATCTGTTGGAGGAGGGC 1140  
 QY 1475 AAGATGACGGCCAGGTCGTGAGTCCGCAAGCCATTCTCGCGGCAAGAGTGCGC 1534  
 Db 1141 AAGATGACGGCCAGGTCGTAGAGAGCGCAAGCCATCTGCGGGAAGCAAGTGCGC 1200  
 QY 1535 GTGACCAAAAGTCAAGTCTCCGCGGATGACCCCAACCCCGTATGTCACCTCC 1594  
 Db 1201 GTGACCAAAAGTCAAGTCTCCGCGGATGACCCCAACCCCGTATGTCACCTCC 1260  
 QY 1595 AACACCAACATGTGCGCGGTGATTGACGGGAACACCAACCTTCGAGCAACGAGCCG 1654  
 Db 1261 AACACCAACATGTGCGCGGTGATTGACGGGAACCTTCGAGCAACGAGCCG 1320  
 QY 1655 TTGACGAGACCGGATGTTCAATTTGAATCAACCCCGCTCTGAGCATGACTTTGGCAAG 1714  
 Db 1321 CTCCAGGACCGGATGTTCAAGTTCAAGCTCAACCAAGCGCTGAGCAAGACTTTGGCAAG 1380  
 QY 1715 GTGACAAAGCAGGAAGTCAAGAGTCTTCCGCTGGGCGCAGATCAAGTGAACGAGGTG 1774  
 Db 1381 GTGACAAAGCAGGAAGTCAAGAGTCTTCCGCTGGGCGTCAAGTCAAGTGAACGAGGTG 1440  
 QY 1775 GCGCATGAGTTCTACGTCAGAAAGGGTGAGGCCAACAAGACCCGCGCGATGACGCG 1834  
 Db 1441 ACTCAGAGTTTACGTCAGAAAGGGTGAGGTAAGAAAGAGCGCGCGCAATGACGCA 1500  
 QY 1835 GATAAAAGCAGCCCAAGCGGCTGCGCTCAGTCCGATCCATGACGTCAGACGCG 1894  
 Db 1501 GATATAAGTGAAGCCCAAGCGGCTGTCGTCAGTTGGCAGCCATGACGTCAGACGCG 1560  
 QY 1895 GAAGAGCTCCGCTGACTTTGCCGACAGGT 1925  
 Db 1561 GA--AGCTCCGCTGACTACGCGGACAGAT 1588

Search completed: July 20, 2003, 13:23:48  
 Job time : 234.626 secs





Feature	Start	End	Strand	Label	Description
Promoter	236..299		+	P5_promoter	
Protein binding site	237..245		+	YY1 factor	
TATA box	270..275		+	P5_TATA-Box	
Miscellaneous feature	299..306		+	YY1/p5 RNA	
CDS	335..2206		+	Rep 78	
CDS	335..2272		+	Rep 68	
Intron	1924..2220		+	Rep 68 and Rep 40	
TATA box	857..862		+	P19_TATA_Box	
Miscellaneous feature	882..883		+	P19 RNA	
CDS	1007..2206		+	Rep 52	
CDS	1007..2272		+	Rep 40	
TATA box	1836..1841		+	P40_TATA-BOX	
Miscellaneous feature	1875..1876		+	P40 RNA	
CDS	2223..4433		+	VP1 protein	
CDS	2634..4433		+	VP2 protein	
CDS	2829..4433		+	VP3 protein	
polyA signal	4447..4452		+	Capsid protein	
repeat unit	4576..4718		+	VP4 protein	
ITR	4718..4718		+	VP5 protein	
ITR	4718..4718		+	VP6 protein	
ITR	4718..4718		+	VP7 protein	
ITR	4718..4718		+	VP8 protein	
ITR	4718..4718		+	VP9 protein	
ITR	4718..4718		+	VP10 protein	
ITR	4718..4718		+	VP11 protein	
ITR	4718..4718		+	VP12 protein	
ITR	4718..4718		+	VP13 protein	
ITR	4718..4718		+	VP14 protein	
ITR	4718..4718		+	VP15 protein	
ITR	4718..4718		+	VP16 protein	
ITR	4718..4718		+	VP17 protein	
ITR	4718..4718		+	VP18 protein	
ITR	4718..4718		+	VP19 protein	
ITR	4718..4718		+	VP20 protein	
ITR	4718..4718		+	VP21 protein	
ITR	4718..4718		+	VP22 protein	
ITR	4718..4718		+	VP23 protein	
ITR	4718..4718		+	VP24 protein	
ITR	4718..4718		+	VP25 protein	
ITR	4718..4718		+	VP26 protein	
ITR	4718..4718		+	VP27 protein	
ITR	4718..4718		+	VP28 protein	
ITR	4718..4718		+	VP29 protein	
ITR	4718..4718		+	VP30 protein	
ITR	4718..4718		+	VP31 protein	
ITR	4718..4718		+	VP32 protein	
ITR	4718..4718		+	VP33 protein	
ITR	4718..4718		+	VP34 protein	
ITR	4718..4718		+	VP35 protein	
ITR	4718..4718		+	VP36 protein	
ITR	4718..4718		+	VP37 protein	
ITR	4718..4718		+	VP38 protein	
ITR	4718..4718		+	VP39 protein	
ITR	4718..4718		+	VP40 protein	
ITR	4718..4718		+	VP41 protein	
ITR	4718..4718		+	VP42 protein	
ITR	4718..4718		+	VP43 protein	
ITR	4718..4718		+	VP44 protein	
ITR	4718..4718		+	VP45 protein	
ITR	4718..4718		+	VP46 protein	
ITR	4718..4718		+	VP47 protein	
ITR	4718..4718		+	VP48 protein	
ITR	4718..4718		+	VP49 protein	
ITR	4718..4718		+	VP50 protein	
ITR	4718..4718		+	VP51 protein	
ITR	4718..4718		+	VP52 protein	
ITR	4718..4718		+	VP53 protein	
ITR	4718..4718		+	VP54 protein	
ITR	4718..4718		+	VP55 protein	
ITR	4718..4718		+	VP56 protein	
ITR	4718..4718		+	VP57 protein	
ITR	4718..4718		+	VP58 protein	
ITR	4718..4				

[illegible]

QY 541 CGTAGTAGGCCCCGGAGGCCCTCTTCTTGTTCAGTTCGAGAGGGCGAGTCTCTACTT 600  
DB 541 CGTAGTAGGCCCCGGAGGCCCTCTTCTTGTTCAGTTCGAGAGGGCGAGTCTCTACTT 600  
QY 601 CCACCTCCATATTTCTGTGTGAGACCACGGGGGTCAAATCCATGTGTGGGCCCTTCTT 660  
DB 601 CCACCTCCATATTTCTGTGTGAGACCACGGGGGTCAAATCCATGTGTGGGCCCTTCTT 660  
QY 661 GAGTCAGATTAGGGACAAGCTGTGTGAGACCATCTACCGCGGGATCGAGCCGACTGCC 720  
DB 661 GAGTCAGATTAGGGACAAGCTGTGTGAGACCATCTACCGCGGGATCGAGCCGACTGCC 720  
QY 721 CAACCTGTTGCGCGGTGACCAAGACGCGTAAATGCGCCGAGGGGGGAACAAGTGTGTA 780  
DB 721 CAACCTGTTGCGCGGTGACCAAGACGCGTAAATGCGCCGAGGGGGGAACAAGTGTGTA 780  
QY 781 CGAGTGTACATCCCCCACTACCTCCTGCCAAGACTCAGCCCGAGCTGCAGTGGCGTG 840  
DB 781 CGAGTGTACATCCCCCACTACCTCCTGCCAAGACTCAGCCCGAGCTGCAGTGGCGTG 840  
QY 841 GACTAACATGAGAGAGTATATAGCGCTGTGTAACCTGCGCGGCAAAACGGCTCGT 900  
DB 841 GACTAACATGAGAGAGTATATAGCGCTGTGTAACCTGCGCGGCAAAACGGCTCGT 900  
QY 901 GCGCGAGCACTGACCCACGTCAGCCAGACCCAGAGCAGAAACAAGAAATCTGAACCC 960  
DB 901 GCGCGAGCACTGACCCACGTCAGCCAGACCCAGAGCAGAAACAAGAAATCTGAACCC 960  
QY 961 CAATTCTGACGCGCTGTCTATCCGTCAAAAAATCTCCGCGCGCTACATGGAAGTGTGCG 1020  
DB 961 CAATTCTGACGCGCTGTCTATCCGTCAAAAAATCTCCGCGCGCTACATGGAAGTGTGCG 1020  
QY 1021 GTGCTGTGTGACCGGGGCATCACCTCCGAGAAAGCATGTGATCCAGAGAACAGGCTC 1080  
DB 1021 GTGCTGTGTGACCGGGGCATCACCTCCGAGAAAGCATGTGATCCAGAGAACAGGCTC 1080  
QY 1081 GTACATCTCTTCAACGCGCTTCCAACTGCGGGTCCAGATCAAGGCCGCTCTGACAA 1140  
DB 1081 GTACATCTCTTCAACGCGCTTCCAACTGCGGGTCCAGATCAAGGCCGCTCTGACAA 1140  
QY 1141 TGCCGGCAAGATCATGGCGCTGACCAATTCGCGCCGACTACCTGTGAGGCCCGCTCC 1200  
DB 1141 TGCCGGCAAGATCATGGCGCTGACCAATTCGCGCCGACTACCTGTGAGGCCCGCTCC 1200  
QY 1201 GCGCGGACATTAATAAACCAACCGCATCTACCTGTGAGCTGAACGGCTACGAACC 1260  
DB 1201 GCGCGGACATTAATAAACCAACCGCATCTACCTGTGAGCTGAACGGCTACGAACC 1260  
QY 1261 TGCCCTACGCGGCTCCGCTTTCTCGGCTGGGCCAGAAAAAGTTCCGGGAAGCGCAAC 1320  
DB 1261 TGCCCTACGCGGCTCCGCTTTCTCGGCTGGGCCAGAAAAAGTTCCGGGAAGCGCAAC 1320  
QY 1321 CATCTGGCTTTTGGGCCCGGCAACCGGCAAGAACCAACATCGCGGAAGCCATCGCCA 1380  
DB 1321 CATCTGGCTTTTGGGCCCGGCAACCGGCAAGAACCAACATCGCGGAAGCCATCGCCA 1380  
QY 1381 CGCCGTGCTTCTACGCGCTGCTCACTGAGCAATGAGAACTTCCCTCAATGATTG 1440  
DB 1381 CGCCGTGCTTCTACGCGCTGCTCACTGAGCAATGAGAACTTCCCTCAATGATTG 1440  
QY 1441 CGTCGACAAAGATGTGATCTGTGTGAGAGGGCAAGATGACGCGCAAGTCTGTGAGTC 1500  
DB 1441 CGTCGACAAAGATGTGATCTGTGTGAGAGGGCAAGATGACGCGCAAGTCTGTGAGTC 1500  
QY 1501 CGCCAAAGGCATCTCGCGCGCAGCAAGGTGCGGTGACCAAAAAGTGAAGTCTCGC 1560  
DB 1501 CGCCAAAGGCATCTCGCGCGCAGCAAGGTGCGGTGACCAAAAAGTGAAGTCTCGC 1560  
QY 1561 CCAGATCGACCCCAACCCCGGTGATCTCTCACTTCCAAACCAACATGTGCGCGTGAATTGA 1620  
DB 1561 CCAGATCGACCCCAACCCCGGTGATCTCTCACTTCCAAACCAACATGTGCGCGTGAATTGA 1620

QY 1621 CGGGAACAGCACCACTTCAGAGCACAGCAGCCGTTGACAGACCGGATGTTCAAATTTGA 1680  
DB 1621 CGGGAACAGCACCACTTCAGAGCACAGCAGCCGTTGACAGACCGGATGTTCAAATTTGA 1680  
QY 1681 ACTGACCCCGCTGTGAGCATGACTTTGGCAAGGTGACAAAGCAGAAAGTCAAAGAGTT 1740  
DB 1681 ACTGACCCCGCTGTGAGCATGACTTTGGCAAGGTGACAAAGCAGAAAGTCAAAGAGTT 1740  
QY 1741 CTTCCGCTGGGCGCAGATCAAGTACCCGAGTGGCGCATGAGTTCTACGTCAAGAAAGG 1800  
DB 1741 CTTCCGCTGGGCGCAGATCAAGTACCCGAGTGGCGCATGAGTTCTACGTCAAGAAAGG 1800  
QY 1801 TGGAGCCAACAAAAGACCCCGCCCGATGACGCGGATTAAGCCGAGCCCAAGCGGCTG 1860  
DB 1801 TGGAGCCAACAAAAGACCCCGCCCGATGACGCGGATTAAGCCGAGCCCAAGCGGCTG 1860  
QY 1861 CCCCTCAGTCGCGGATCCATGACGTGACAGCGCGAAGAGCTCCGGTGAATTTGCCGA 1920  
DB 1861 CCCCTCAGTCGCGGATCCATGACGTGACAGCGCGAAGAGCTCCGGTGAATTTGCCGA 1920  
QY 1921 CAGGTACCAAAAACAAATGTTCTGTCACGCGGGCATGCTTCAGATGCTGTTCCCTGCA 1980  
DB 1921 CAGGTACCAAAAACAAATGTTCTGTCACGCGGGCATGCTTCAGATGCTGTTCCCTGCA 1980  
QY 1981 GACATGCGAGAGATGAATCAGAAATTTCAACATTTGCTTCAACGCGGTCAGAAAAGAGCGTATCG 2040  
DB 1981 GACATGCGAGAGATGAATCAGAAATTTCAACATTTGCTTCAACGCGGTCAGAAAAGAGCGTATCG 2040  
QY 2041 TTTACAGTGTCTTCCCGCGGTTCAGAAATCTCAACCGGTCTGAGAAAAGAGCGTATCG 2100  
DB 2041 TTTACAGTGTCTTCCCGCGGTTCAGAAATCTCAACCGGTCTGAGAAAAGAGCGTATCG 2100  
QY 2101 GAAACTCTGTGCCATTCAATCTGCTGGGGCGGGCTCCCGAGATTGCTTGTGCGGCTG 2160  
DB 2101 GAAACTCTGTGCCATTCAATCTGCTGGGGCGGGCTCCCGAGATTGCTTGTGCGGCTG 2160  
QY 2161 CGATCTGTCAACGTGACCTGATGACTGTGTTCTGAGCAATTAATGACTTAAACAG 2220  
DB 2161 CGATCTGTCAACGTGACCTGATGACTGTGTTCTGAGCAATTAATGACTTAAACAG 2220  
QY 2221 GTATGCTGCGGATGTTATCTTCCAGATTGGCTCGAGCAACCTCTCTGAGGCGATTC 2280  
DB 2221 GTATGCTGCGGATGTTATCTTCCAGATTGGCTCGAGCAACCTCTCTGAGGCGATTC 2280  
QY 2281 GCGAGTGTGGGACTTGAAACCTGAGACCCCGAAGCCCAAGCCAAACGAGCAAG 2340  
DB 2281 GCGAGTGTGGGACTTGAAACCTGAGACCCCGAAGCCCAAGCCAAACGAGCAAG 2340  
QY 2341 ACGACGCGGGGTCTGTGCTTCTGCTACAAAGTACTTGGAAACCTTCAACGAGCTCG 2400  
DB 2341 ACGACGCGGGGTCTGTGCTTCTGCTACAAAGTACTTGGAAACCTTCAACGAGCTCG 2400  
QY 2401 ACAAGGGGAGACCCGTCACGCGCGGACGAGCGGCGCTCGAGCAGCAAGAGGCTTACG 2460  
DB 2401 ACAAGGGGAGACCCGTCACGCGCGGACGAGCGGCGCTCGAGCAGCAAGAGGCTTACG 2460  
QY 2461 ACCAGAGCTCAAAAGCGGTGACATCCGTACCTGCGGTATAACACGCGGAGCGAGT 2520  
DB 2461 ACCAGAGCTCAAAAGCGGTGACATCCGTACCTGCGGTATAACACGCGGAGCGAGT 2520  
QY 2521 TTTACAGAGCTGTGCAAGAAATACGTCTTTTGGGGGCAACCTCGGGCGAGCATCTCC 2580  
DB 2521 TTTACAGAGCTGTGCAAGAAATACGTCTTTTGGGGGCAACCTCGGGCGAGCATCTCC 2580  
QY 2581 AGGCAAGAAAGCGGTTCTCGAACTCTCGGTCTGTGTTGAGGAAAGCGCTTAAGACGGCTC 2640  
DB 2581 AGGCAAGAAAGCGGTTCTCGAACTCTCGGTCTGTGTTGAGGAAAGCGCTTAAGACGGCTC 2640  
QY 2641 CTGGAAGAAACGTCCGTTAGAGAGTGCACAAAGAGCCAGACTCTCTCGGCGATCG 2700  
DB 2641 CTGGAAGAAACGTCCGTTAGAGAGTGCACAAAGAGCCAGACTCTCTCGGCGATCG 2700  
QY 2701 GCAAGACAGGCGCAGCAGCCCGCTAATAAAGAGACTCAATTTTGTTCAGACTGGCGACTCAG 2760

```
|||||
Db 2701 GCAAGACAGGCCAGCAGCCCGCTAAAAAGAGACTCAATTTGTCAGACTGCGCACTCAG 2760
QY 2761 AGTCAGTCCCGATCCACAACTCTCGGAGAACCTCCAGCAACCCCGCTGTGTGGAC 2820
Db 2761 AGTCAGTCCCGATCCACAACTCTCGGAGAACCTCCAGCAACCCCGCTGTGTGGAC 2820
QY 2821 CTACTACAATGCTTCAGCGCGGTGGCGCAGCAATGGCAGACAATTAACGAAAGCGCGCAGC 2880
Db 2821 CTACTACAATGCTTCAGCGCGGTGGCGCAGCAATGGCAGACAATTAACGAAAGCGCGCAGC 2880
QY 2881 GAGTGGGTAAATGCTCAGGAAATTTGGCATTTGCAATGCTGCGGCGAGAGTCA 2940
Db 2881 GAGTGGGTAAATGCTCAGGAAATTTGGCATTTGCAATGCTGCGGCGAGAGTCA 2940
QY 2941 TCACCAACAGCAGCCGACCTGGGCTTGCCCACTACAATAACCACTCTACAAGCAA 3000
Db 2941 TCACCAACAGCAGCCGACCTGGGCTTGCCCACTACAATAACCACTCTACAAGCAA 3000
QY 3001 TCTCCAGTGTTCACAGCGGGGGCCAGCAACGACCACTACTTCGGCTACAGACCCCT 3060
Db 3001 TCTCCAGTGTTCACAGCGGGGGCCAGCAACGACCACTACTTCGGCTACAGACCCCT 3060
QY 3061 GGGGGTATTTGATTTCAACAGATTCCAGTCCCACTTTTCAACCACTGAGTGGCAGCGAC 3120
Db 3061 GGGGGTATTTGATTTCAACAGATTCCAGTCCCACTTTTCAACCACTGAGTGGCAGCGAC 3120
QY 3121 TCATCAACAACAATGGGGATTCGGGCCAAGAGACTCAACTTCAACTCTCAACATCC 3180
Db 3121 TCATCAACAACAATGGGGATTCGGGCCAAGAGACTCAACTTCAACTCTCAACATCC 3180
QY 3181 AAGTCAAGAGGTCAAGCAAGATGATGGCTCAACCACTGCTAATTAACCTTACAGCA 3240
Db 3181 AAGTCAAGAGGTCAAGCAAGATGATGGCTCAACCACTGCTAATTAACCTTACAGCA 3240
QY 3241 CGGTTCAAGTCTTCGGGACTCGGAGTACCACTCCGTACGTCCTCGGCTCTGGCACC 3300
Db 3241 CGGTTCAAGTCTTCGGGACTCGGAGTACCACTCCGTACGTCCTCGGCTCTGGCACC 3300
QY 3301 AGGGCTGCTCCCTCCGTTCCCGGCGGACGTTTCATGATTCGCAATACGGCTACCTGA 3360
Db 3301 AGGGCTGCTCCCTCCGTTCCCGGCGGACGTTTCATGATTCGCAATACGGCTACCTGA 3360
QY 3361 CGCTCAACAATGGCAGCCAGCCGTTGGGACGTTTCATCTTTACTGCTGGAATATTCC 3420
Db 3361 CGCTCAACAATGGCAGCCAGCCGTTGGGACGTTTCATCTTTACTGCTGGAATATTCC 3420
QY 3421 CTTCTCAGATGCTGAGAACGGGCAACAATTTCCTTACGCTACACCTTTGAGGAAGTGC 3480
Db 3421 CTTCTCAGATGCTGAGAACGGGCAACAATTTCCTTACGCTACACCTTTGAGGAAGTGC 3480
QY 3481 CTTTCCACAGCAGCTACGCGCAGCAGGAGCTGGAACGGCTGATGAATCCTCTCATCG 3540
Db 3481 CTTTCCACAGCAGCTACGCGCAGCAGGAGCTGGAACGGCTGATGAATCCTCTCATCG 3540
QY 3541 ACCAATACCTGTATTACCTGAACAGAACTCAAAATCAGTCCGGAAGTCCCAAAACAAG 3600
Db 3541 ACCAATACCTGTATTACCTGAACAGAACTCAAAATCAGTCCGGAAGTCCCAAAACAAG 3600
QY 3601 ACTTGCTGTTAGCGGTGGGTCTCCAGCTGGCATGTCTGTTACGCCCAAAACTGGCTAC 3660
Db 3601 ACTTGCTGTTAGCGGTGGGTCTCCAGCTGGCATGTCTGTTACGCCCAAAACTGGCTAC 3660
QY 3661 CTGACCCCTGTTATCGGCAAGCGGCTTTTAAACAAAAACAGACAACAACAGCA 3720
Db 3661 CTGACCCCTGTTATCGGCAAGCGGCTTTTAAACAAAAACAGACAACAACAGCA 3720
QY 3721 ATTTTACCTGAGCTGTGCTTCAAAATATTAATCAATGGGCGTGAATCATCAACC 3780
Db 3721 ATTTTACCTGAGCTGTGCTTCAAAATATTAATCAATGGGCGTGAATCATCAACC 3780
QY 3781 CTGCACTGTATGGCTCACACAAGAGCAGAAAGCAAGTTCTTTCCATGAGCGGTG 3840
|||||
```

```
Db 3781 CTGCACTGCTATGGCCTCACACAAGACGACGAAGACAAAGTTCTTCCATGAGCGGTG 3840
QY 3841 TCATGATTTTGGAAAAAGAGAGCGCCGAGACTTCAAAACACTGCAATGGACATGTATGA 3900
Db 3841 TCATGATTTTGGAAAAAGAGAGCGCCGAGACTTCAAAACACTGCAATGTGATGA 3900
QY 3901 TTACAGCAAGAGGAATTAAGCCACTAACCCCTGTGGCCACCGAAAGATTGGACCG 3960
Db 3901 TTACAGCAAGAGGAATTAAGCCACTAACCCCTGTGGCCACCGAAAGATTGGACCG 3960
QY 3961 TGGCAGTCAATTTCCAGAGCAGCAGACAGACCCTGCGACCGGAGATGTGCATGTATGG 4020
Db 3961 TGGCAGTCAATTTCCAGAGCAGCAGACAGACCCTGCGACCGGAGATGTGCATGTATGG 4020
QY 4021 GAGCATTAACCTGGCATGTGTGGCAAGATAGAGACGTGTACTGCAAGGTCCCATTTGGG 4080
Db 4021 GAGCATTAACCTGGCATGTGTGGCAAGATAGAGACGTGTACTGCAAGGTCCCATTTGGG 4080
QY 4081 CCAAAATCTCTACACAGATGACACTTTCACCCCGTCTCTCTTATGGCGGCTTTGAC 4140
Db 4081 CCAAAATCTCTACACAGATGACACTTTCACCCCGTCTCTCTTATGGCGGCTTTGAC 4140
QY 4141 TCAAGAACCCGCTCTCAGATCCTCATCAAAAAACAGCGCTGTCTGCGAATCTCCGG 4200
Db 4141 TCAAGAACCCGCTCTCAGATCCTCATCAAAAAACAGCGCTGTCTGCGAATCTCCGG 4200
QY 4201 CGAGTTTTCAGCTACAAAGTTGCTTCATTCATCACCCCAATCTCCACAGACAAGTGA 4260
Db 4201 CGAGTTTTCAGCTACAAAGTTGCTTCATTCATCACCCCAATCTCCACAGACAAGTGA 4260
QY 4261 GTGTGAAATTTGAATGGAGCTGCAAGAAAGAAACAGCAAGCGCTGGAATCCCGAAGTGC 4320
Db 4261 GTGTGAAATTTGAATGGAGCTGCAAGAAAGAAACAGCAAGCGCTGGAATCCCGAAGTGC 4320
QY 4321 AGTACATATCCAATTATGCAAAATCTGCCAAGTTGATTTTACTGTGACAAACAATGAC 4380
Db 4321 AGTACATATCCAATTATGCAAAATCTGCCAAGTTGATTTTACTGTGACAAACAATGAC 4380
QY 4381 TTTTACTGAGCCTGCCCCATTTGGCACCCTTACCTTACCCGTCCCTGTAAATTAAGTGC 4440
Db 4381 TTTTACTGAGCCTGCCCCATTTGGCACCCTTACCTTACCCGTCCCTGTAAATTAAGTGC 4440
QY 4441 TTAATCAATAAACCGGTTGATTCGTTTCAAGTTGAACCTTGGTCTCTGTCTCTTATCT 4500
Db 4441 TTAATCAATAAACCGGTTGATTCGTTTCAAGTTGAACCTTGGTCTCTGTCTCTTATCT 4500
QY 4501 TATCGTTTACCATGTTATAGCTTACACATTAACCTGCTTGGTGGCTTGGCATTAAG 4560
Db 4501 TATCGTTTACCATGTTATAGCTTACACATTAACCTGCTTGGTGGCTTGGCATTAAG 4560
QY 4561 ACTTACGTCATCGGGTTACCCCTAGTATGAGTGGAGTTGCCACTCCCTCTCTGCGGCTGCG 4620
Db 4561 ACTTACGTCATCGGGTTACCCCTAGTATGAGTGGAGTTGCCACTCCCTCTCTGCGGCTGCG 4620
QY 4621 TCGCTCGGTGGGCTGCGGACCAAGGTCGCGACAGCGGAGAGCTGTGCTGCGGCGC 4680
Db 4621 TCGCTCGGTGGGCTGCGGACCAAGGTCGCGGACCAAGGTCGCGAGAGAGCTGTGCTGCGGCGC 4680
QY 4681 CCACCGAGCAGCGAGCGCGCAGAGAGGAGTGGGCAA 4718
Db 4681 CCACCGAGCAGCGAGCGCGCAGAGAGGAGTGGGCAA 4718
```

RESULT 2  
AAF23749 standard; DNA; 4683 BP.

AAF23749;

28-MAR-2001 (first entry)

AAV6 DNA sequence.

XX



KW AAV6; gene therapy; AAV viral vector; cystic fibrosis; cancer; AIDS;  
KM atherosclerosis; sickle cell anaemia; thalassaemia;  
KW blood clotting disorder; diabetes; ss.  
OS Adeno associated virus.  
XX US6156303-A.  
FN 05-DEC-2000.  
PD 11-JUN-1997; 97US-0873168.  
XX 11-JUN-1997; 97US-0873168.  
PF 11-JUN-1997; 97US-0873168.  
XX 11-JUN-1997; 97US-0873168.  
PR (UNIW ) UNIV WASHINGTON.  
XX Russell DW, Rutledge EA;  
PI WPI; 2001-060164/07.  
XX Adeno-associated virus serotype 6 and viral vector derived from it for  
PT gene therapy of cystic fibrosis, cancer, acquired immunodeficiency  
PT syndrome, sickle cell anemia, thalassemia and diabetes  
XX Claim 1; Fig 1; 50pp; English.  
PS  
XX The present invention relates to adeno-associated virus serotypes. The  
CC present sequence is the DNA sequence of one such serotype (AAV6). AAV6  
CC can be used to construct AAV viral vectors for use in gene therapy for a  
CC range of disorders: cystic fibrosis, cancer, AIDS, atherosclerosis,  
CC sickle cell anaemia, thalassaemia, blood clotting disorders and diabetes.  
CC The AAV viral vectors have increased transduction efficiency of a  
CC particular host cell as the AAV virion containing the AAV vector genome  
CC can be modified to express a capsid protein of an AAV serotype that  
CC transduces the selected host cell.  
XX  
SQ Sequence 4683 BP; 1114 A; 1363 C; 1277 G; 929 T; 0 other;  
  
Query Match 90.3%; Score 4258; DB 22; Length 4683;  
Best Local Similarity 94.7%; Pred. No. 0;  
Matches 4474; Conservative 0; Mismatches 205; Indels 43; Gaps 5;  
  
QY 1 TTGCCCACTCCCTCTCTGCGCGCTCGCTCGCTGCGGCGCGGACCAAGTCCGC 60  
Db 1 TTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGCGGACCAAGTCCGC 60  
QY 61 AGACGGCAGAGCTCTGCTCTGCGGCGCGGACCGAGCGGCGGAGAGGAGTG 120  
Db 61 CGACGCCGGGCTTTGCGCGGCGGCTCACTGAGCGAGCGGCGGAGAGGAGTG 120  
QY 121 GGCACTCCATCACTAGGGGTAATCGGAAGCGCCTCCACGCTGCGCTCAGCGCTGA 180  
Db 121 GCCAATCCATCACTAGGGG-----TCCTGAGGGGTGAGTCTGA 163  
QY 181 CGTAATAAGTCATAGGG---GAGTGGTCTGTATTAGCTGTCACTGAGTCTTTGC 237  
Db 164 CGTAATAAGTCATAGGGTGAAGGTCCCTGTATTAGAGTCACTGAGTG-TTTTGC 222  
QY 238 GACATTTTGGACACACAGTGGCCATTAGGGTATATATGCGGAGTGAGCGAGAT 297  
Db 223 GACATTTTGGACACACATGTGTGTCACGCTGGGTATTTAAGCCGAGTGAGCAGCGGT 282  
QY 298 CTCCATTTTGAACCGGAATTGAACGACGACGACCATGCGGCGCTTCTACGAGATCGT 357  
Db 283 CTCCATTTTGAAGCGGAGGTTGAACGCGACGCGCATGCGGCGTTTACGAGATTGT 342  
QY 358 GATCAAGTGGCGGAGCGACCTGAGCAGACCTGCGGCGATTCTGACTGTTGTGAG 417  
Db 343 GATTAAAGTCCCGACGACCTTGAAGCATCTGCCGCGCATTTCTGACAGCTTTGTGA 402  
QY 418 CTGGTGGCGGAGGAAGATGGAGCTGCGCCCGGATTCTGACATGATCTGAATCTGAT 477  
Db 403 CTGGTGGCGGAGGAAGATGGAGTTCGCCGCGCATTTCTGACATGATCTGAATCTGAT 462

QY 478 TGACAGGCACCCCTGACCCGTGGCCGAGAAGCTGCAGCGCACTTCTGTCCATGCGG 537  
Db 463 TGACAGGCACCCCTGACCCGTGGCCGAGAAGCTGCAGCGCACTTCTGTCCATGCGG 522  
QY 538 CCGCGTGAAGTAAAGCCCCGGAGGCCCTCTTCTTTGTTCACTTCAGAAAGGCGAGTCTTA 597  
Db 523 CCGCGTGAAGTAAAGCCCCGGAGGCCCTCTTCTTTGTTCACTTCAGAAAGGCGAGTCTTA 582  
QY 598 CTTCACCTCCATATTCTGTGAGAGACCAAGGGGTCAATTCATGTGTGCGCGCTT 657  
Db 583 CTTCACCTCCATATTCTGTGAGAGACCAAGGGGTCAATTCATGTGTGCGCGCTT 642  
QY 658 CCTGAGTCAGATTAGGAGCAAGCTGTGAGACCATCTACCGCGGATCGAGCGACCTT 717  
Db 643 CCTGAGTCAGATTAGGAGCAAGCTGTGAGACCATCTACCGCGGATCGAGCGACCTT 702  
QY 718 GCCCAACTGTTGCGCGTGAACCAAGCGCTAATGCGCGGAGGGGGAACAAGTGTGT 777  
Db 703 GCCCAACTGTTGCGCGTGAACCAAGCGCTAATGCGCGGAGGGGGAACAAGTGTGT 762  
QY 778 GAGCAGTGTCTACATCCCACTAATCTCTGCGGCAAGCTCAGCCGAGCTGAGTGGGC 837  
Db 763 GAGCAGTGTCTACATCCCACTAATCTCTGCGGCAAGCTCAGCCGAGCTGAGTGGGC 822  
QY 838 GTGACTAATCATGAGAGATATATAAGCGCTGTTGAATCTGCGCGCAACGCGCT 897  
Db 823 GTGACTAATCATGAGAGATATATAAGCGCTGTTGAATCTGCGCGCAACGCGCT 882  
QY 898 CGTGGCGCAGCACCCTGACCCAGTCAAGCCAGACCCAGAGCAAGAAAGAAATCTGAA 957  
Db 883 CGTGGCGCAGCACCCTGACCCAGTCAAGCCAGACCCAGAGCAAGAAAGAAATCTGAA 942  
QY 958 CCCCAATTCTGACGCGCTGTCTATCCGCTCAAAAACCTCCGCGCTACATGAGCTGT 1017  
Db 943 CCCCAATTCTGACGCGCTGTCTATCCGCTCAAAAACCTCCGCGCTACATGAGCTGT 1002  
QY 1018 CCGGTGGCTGTGTGACCGGGGCACTACCTCCGAGAGAGTGGATCCAGAGACCAAGC 1077  
Db 1003 CCGGTGGCTGTGTGACCGGGGCACTACCTCCGAGAGAGTGGATCCAGAGACCAAGC 1062  
QY 1078 CTGCTACATCTCTTCAACGCGCTTCAACTCGCGGTCCAGATCAAGGCGCTGTGA 1137  
Db 1063 CTGCTACATCTCTTCAACGCGCTTCAACTCGCGGTCCAGATCAAGGCGCTGTGA 1122  
QY 1138 CAATGCCGGAAGATCATGGCGCTGACCAATCCGCGCGCACTACCTGTAGGCCCGC 1197  
Db 1123 CAATGCCGGAAGATCATGGCGCTGACCAATCCGCGCGCACTACCTGTAGGCCCGC 1182  
QY 1198 TCCGCGCGCGGACATTAAACCAACCGCATCTACCGCATCTCTGAGCTGAACGGTACGA 1257  
Db 1183 TCCGCGCGCGGACATTAAACCAACCGCATCTACCGCATCTCTGAGCTGAACGGTACGA 1242  
QY 1258 ACTGCTTACGCGCGCTCCGTCTTCTCGGCTGGGCGCAGAAAAGTTGGGAGGCGCA 1317  
Db 1243 CCTGCTTACGCGCGCTCCGTCTTCTCGGCTGGGCGCAGAAAAGTTGGGAGGCGCA 1302  
QY 1318 CACCATCTGCTGTTTGGCGCGGCAACCAAGGCAAGCAATCGCGGAAGCCATCGC 1377  
Db 1303 CACCATCTGCTGTTTGGCGCGGCAACCAAGGCAAGCAATCGCGGAAGCCATCGC 1362  
QY 1378 CCAAGCGGTGCGCTTCTACGCGCTGCGTCAACTGAGCAATGAGAACTTCCCTTCAATGA 1437  
Db 1363 CCAAGCGGTGCGCTTCTACGCGCTGCGTCAACTGAGCAATGAGAACTTCCCTTCAATGA 1422  
QY 1438 TTGCGTGCACAAGATGTGATCTGCTGGAGAGGCAAGATGACGCGCAAGTCTGGA 1497  
Db 1423 TTGCGTGCACAAGATGTGATCTGCTGGAGAGGCAAGATGACGCGCAAGTCTGGA 1482  
QY 1498 GTCCGCCAAGGCCATTCTCGCGCGGCAAGAGTGGCGGTGACCAAAAAGTGCAGTGTCTC 1557  
Db 1483 GTCCGCCAAGGCCATTCTCGCGCGGCAAGAGTGGCGGTGACCAAAAAGTGCAGTGTCTC 1542

QY	1558	CGCCCAAGATCGACCCCAACCCCGGTGATCGTCACTTCCAAACCAACATGTGCGCCGTGAT	1617
Db	1543	CGCCCAAGATCGATCCCAACCCCGGTGATCGTCACTTCCAAACCAACATGTGCGCCGTGAT	1602
QY	1618	TGACGGGAACAGCACCACTTTCAGACCAACAGCAGCCGTTGCAAGACCAGATGTTCAAATT	1677
Db	1603	TGACGGGAACAGCACCACTTTCAGACCAACAGCAGCCGTTGCAAGACCAGATGTTCAAATT	1662
QY	1678	TGAACCTCACCCCGCTCTGAGCATGACTTTGGCAAGGTGCAAAAGCAGGAAGTCAAGA	1737
Db	1663	TGAACCTCACCCCGCTCTGAGCATGACTTTGGCAAGGTGCAAAAGCAGGAAGTCAAGA	1722
QY	1738	GTTCTTCCGCTGGGCGCAGGATCACTGACCGAGGTGGCGCATGAGTTCTACGTCAAGAA	1797
Db	1723	GTTCTTCCGCTGGGCGCAGGATCACTGACCGAGGTGGCGCATGAGTTCTACGTCAAGAA	1782
QY	1798	GGGTGGAGCCCAACAAAAGACCCCGCCCGATGACCGCGGATTAAGCGAGCCCAAGCGGGC	1857
Db	1783	GGGTGGAGCCCAACAAAGAGACCCCGCCCGATGACCGCGGATTAAGCGAGCCCAAGCGGGC	1842
QY	1858	CTGCCCCCTCAGTCGCGGATCCATCGACGTGACAGCGCGGAAGAGCTCCGGTGAATTGGC	1917
Db	1843	CTGCCCCCTCAGTCGCGGATCCATCGACGTGACAGCGCGGAAGAGCTCCGGTGAATTGGC	1902
QY	1918	CGACAGGTACCAAAACAATGTTCTCGTCAACGCGGCGATGCTTCAGATGCTGTTCCCTG	1977
Db	1903	CGACAGGTACCAAAACAATGTTCTCGTCAACGCGGCGATGCTTCAGATGCTGTTCCCTG	1962
QY	1978	CAAGACATGCGAGAGATGAATCAGAATTTCAAATTGCTTCAAGCACGGACGAGAGAGA	2037
Db	1963	CAAAACATGCGAGAGATGAATCAGAATTTCAAATTGCTTCAAGCACGGACGAGAGAGA	2022
QY	2038	CTGTTTCAAGAGTCTTCCCCCGCGCTGTCAGAATCTCAACCCGTCGTCAAGAAAGAGACGTA	2097
Db	2023	CTGTTTCAAGAGTCTTCCCCCGCGCTGTCAGAATCTCAACCCGTCGTCAAGAAAGAGAGTA	2082
QY	2098	TGCGAAACTCTGTGCCATTTCATCTGCTGGGGCGGGCTCCGAGATTGCTTGCTCGGC	2157
Db	2083	TGCGAAACTCTGTGCCATTTCATCTGCTGGGGCGGGCTCCGAGATTGCTTGCTCGGC	2142
QY	2158	CTGCGATCTGCTCAACGTGACCTGAGTGACTGTGTTCTGAGCAATTAATGACTTTAAAC	2217
Db	2143	CTGCGATCTGCTCAACGTGAGTCTGAGTGACTGTGTTCTGAGCAATTAATGACTTTAAAC	2202
QY	2218	CAGGTATGGCTGCCGATGTTATCTTCCAGATTGGCTCGAGCAAACTCTCTGAGGGCA	2277
Db	2203	CAGGTATGGCTGCCGATGTTATCTTCCAGATTGGCTCGAGCAAACTCTCTGAGGGCA	2262
QY	2278	TTCCGAGTGGTGGGACTTGAAACTGGAAGCCCCGAAGCCCAAGCCAAACAGCAAAAGC	2337
Db	2263	TTCCGAGTGGTGGGACTTGAAACTGGAAGCCCCGAAGCCCAAGCCCAAGCAAAAGC	2322
QY	2338	AGGACGACGGCCGGGGTCTGCTGCTTCTTGCTACAAGTACCTCGGAACCTTCAACGGAC	2397
Db	2323	AGGACGACGGCCGGGGTCTGCTGCTTCTTGCTACAAGTACCTCGGAACCTTCAACGGAC	2382
QY	2398	TCGACAAAGGGGGAGCCCGTCAACGCGGCGGACGACGAGCGGCCCTTGAAGCAGCAAGGCT	2457
Db	2383	TCGACAAAGGGGGAGCCCGTCAACGCGGCGGATGCAAGCGGCCCTTGAAGCAGCAAGGCT	2442
QY	2458	ACGACCAAGCAAGCTCAAAAGCGGGTGAATCCGTACCTGCGGTATTAACCAACGCCAGCGCG	2517
Db	2443	ACGACCAAGCAAGCTCAAAAGCGGGTGAATCCGTACCTGCGGTATTAACCAACGCCAGCGCG	2502
QY	2518	AGTTTCAAGGAGCGTCTGCAAGAAGTACGTCTTTTGGGGGCAACTCGGGCGAGCAGTCT	2577
Db	2503	AGTTTCAAGGAGCGTCTGCAAGAAGTACGTCTTTTGGGGGCAACTCGGGCGAGCAGTCT	2562
QY	2578	TCCAGGCCAAGAAGCGGGTCTCGAACCTCTCGGTCTGTTGAAGGAAGCGCTAAGACGG	2637
Db	2563	TCCAGGCCAAGAAGCGGGTCTCGAACCTTTTGGTCTGTTGAAGGAAGGTGCTAAGACGG	2622
QY	2638	CTCCTGGAAAAGAAACGTCCCGGTAGAGCAGTCGCCACAAGACCGACGACTCCTCCTCGGGCA	2697

Db	2623	CTCCTGAAAGAAAGCTCCGGTAGAGCAGTCGCCACAAGAGCCAGACTCTCTCCGGGCA	2682
QY	2698	TCGGCAAGACAGGCCAGCAGCCCCGCTAAAAAGAGACTCAATTTTGTGACACTGGCGACT	2757
Db	2683	TTGGCAAGACAGGCCAGCAGCCCCGCTAAAAAGAGACTCAATTTTGTGACACTGGCGACT	2742
QY	2758	CAGAGTCAGTCCCCGATCCACAACCTCTCGAGAACTCCAGCAACCCCCGCTGTGTGG	2817
Db	2743	CAGAGTCAGTCCCCGATCCACAACCTCTCGAGAACTCCAGCAACCCCCGCTGTGTGG	2802
QY	2818	GACCTACTACAATGGCTTTCAGGCGGTGGCGCACCAATGGCAGACAATAACGAAGCGCCG	2877
Db	2803	GACCTACTACAATGGCTTTCAGGCGGTGGCGCACCAATGGCAGACAATAACGAAGCGCCG	2862
QY	2878	ACGGAGTGGGTAAATGCTCAGGAAATTGGCATTTGGCATTTCCACATGGCTGGGCGCAGAG	2937
Db	2863	ACGGAGTGGGTAAATGCTCAGGAAATTGGCATTTGGCATTTCCACATGGCTGGGCGCAGAG	2922
QY	2938	TCATCACCAACCAAGCAGCCCGCAGCCTGGGCTTGCCCACTTCAATTAACCACTCTACAAGC	2997
Db	2923	TCATCACCAACCAAGCAGCCCGCAGCCTGGGCTTGCCCACTTCAATTAACCACTCTACAAGC	2982
QY	2998	AAATCTCCAGTGTCTCAACGGGGGCCAGCAAGCAACCACTACTTCCGCTACAGCACC	3057
Db	2983	AAATCTCCAGTGTCTCAACGGGGGCCAGCAAGCAACCACTACTTCCGCTACAGCACC	3042
QY	3058	CCTGGGGGTATTTTGATTTCAACAGATTTCCACTGCCACTTTTCCACCAGCTGAGCGCAGC	3117
Db	3043	CCTGGGGGTATTTTGATTTCAACAGATTTCCACTGCCACTTTTCCACCAGCTGAGCGCAGC	3102
QY	3118	GACTCATCAACAACAATTTGGGGATTCCGGCCCCAAGAGACTCAACTTCAAACTTTCAACA	3177
Db	3103	GACTCATCAACAACAATTTGGGGATTCCGGCCCCAAGAGACTCAACTTCAAGCTTTCAACA	3162
QY	3178	TCCAAGTCAAGAGAGTCAAGCAAGATGATGGCGTCACAACCATCGCTAATAACTTTACCA	3237
Db	3163	TCCAAGTCAAGAGAGTCAAGCAAGATGATGGCGTCAAGCAACCATCGCTAATAACTTTACCA	3222
QY	3238	GCACGGTTCAAGTCTTCTCGGACTCGGAGTACCAAGCTTCCGTACGTCCTCGGCTGTGCGC	3297
Db	3223	GCACGGTTCAAGTCTTCTCGGACTCGGAGTACCAAGTTCGTCGTCCTCGGCTGTGCGC	3282
QY	3298	ACCAGGGGCTGCTCCCTCCGTTCCCGCGCAGCGTTCATGATTCGCAATAACGGCTACC	3357
Db	3283	ACCAGGGGCTGCTCCCTCCGTTCCCGCGCAGCGTTCATGATTCGCAATAACGGCTACC	3342
QY	3358	TGACGCTCAACAATGGCAGCCCAAGCCGTGGGACGTTTCATCTTTTACTGCTGGAATATT	3417
Db	3343	TGACGCTCAACAATGGCAGCCCAAGCCGTGGGACGTTTCATCTTTTACTGCTGGAATATT	3402
QY	3418	TCCCTTCTCAGATGCTTGAGAACGGGGCAACAATTACCTTCAAGTACAACCTTTGAGGAAG	3477
Db	3403	TCCCTTCTCAGATGCTTGAGAACGGGGCAACAATTACCTTCAAGTACAACCTTTGAGGAAG	3462
QY	3478	TGCTTTTCCACAGCAGCTACGCGCAGCAGCCAGAGCTGGAACGGCTGATGAATCTCTCA	3537
Db	3463	TGCTTTTCCACAGCAGCTACGCGCAGCAGCCAGAGCTGGAACGGCTGATGAATCTCTCA	3522
QY	3538	TCGACCAATACCTGTATTAACCTGAACAGAACTCAAAATCACTCCGGAAGTGCCCAAAACA	3597
Db	3523	TCGACCAATACCTGTATTAACCTGAACAGAACTCAAAATCACTCCGGAAGTGCCCAAAACA	3582
QY	3598	AGGACTTGTCTGTTAGCCGTGGGTCTCCAGCTGGCATGTCTGTTCAAGCCCAAAAACCTGGC	3657
Db	3583	AGGACTTGTCTGTTAGCCGTGGGTCTCCAGCTGGCATGTCTGTTCAAGCCCAAAAACCTGGC	3642
QY	3658	TACCTTGACCCCTGTTATCGGCAGCAGCGCGTTTCTTAAACAAAAACAGACAACAACA	3717
Db	3643	TACCTTGACCCCTGTTATCGGCAGCAGCGCGTTTCTTAAACAAAAACAGACAACAACA	3702
QY	3718	GCAATTTTACCTGACTGTGTCTTCAAAATATAACCTCAATGGCGGTGAATTCATCATCA	3777

Db	3703	GCAACTTTACCTGGACTGGTGGCTTCAAAATATTAACCTTAATGGGCGTGAATCTATTAATCA	3762
QY	3778	ACCCTGGCACTGCTATGGCCTCACACAAGAAGCAGCAACAAGTTCTTTCCCATGAGCG	3837
Db	3763	ACCCCTGGCACTGCTATGGCCTCACACAAGAAGCAGCAACAAGTTCTTTCCCATGAGCG	3822
QY	3838	GTGTCATGATTTTGGAAAGAGAGCGCCGAGCTTCAACACTGCATTGGACAATGTCA	3897
Db	3823	GTGTCATGATTTTGGAAAGAGAGCGCCGAGCTTCAACACTGCATTGGACAATGTCA	3882
QY	3898	TGATTAACAGACGAAGAGAAATTAAGCCACTAACCCCTGTGGCCACCGAAGATTGGGA	3957
Db	3883	TGATTAACAGACGAAGAGAAATTAAGCCACTAACCCCTGTGGCCACCGAAGATTGGGA	3942
QY	3958	CCGTGGCAGTCAATTTCCAGACGACGACACAGACCCCTGGCAACCGAGATGTGCATGCTA	4017
Db	3943	CTGTGGCAGTCAATTTCCAGACGACGACACAGACCCCTGGCAACCGAGATGTGCATGCTA	4002
QY	4018	TGGAGCATTACCTGGCATGTGTGGCAAGATAGAGCTGTACCTGCAGGGTCCATT	4077
Db	4003	TGGAGCCTTACCTGGAATGTGTGGCAAGACAGACGATTAACCTGCAGGGTCCATT	4062
QY	4078	GGGCCAAAATTCCTCACACAGATGGAACACTTACCCGCTCTCCTTATGGCGGCTTG	4137
Db	4063	GGGCCAAAATTCCTCACACGAGATGGACACTTACCCGCTCTCCTCATGGCGGCTTG	4122
QY	4138	GACTCAAGAACCCTCTCAGATCCTCATCAAAAACACGCCCTGTTCTCGCAATCCTC	4197
Db	4123	GACTTAAGCACCCGCTCTCAGATCCTCATCAAAAACACGCCCTGTTCTCGCAATCCTC	4182
QY	4198	CGCGGAGTTTTCAGCTACAAAGTTTGCTTCAATTCACCAATACTCCACAGGACAAG	4257
Db	4183	CGCGAGAGTTTTCGCTACAAAGTTTGCTTCAATTCACCAATATTCCACAGGACAAG	4242
QY	4258	TGAGTGTGGAATTTGAATGGAGCTGCAGAAAAGAAAACGAACGCTGGAATCCGAAG	4317
Db	4243	TGAGCGTGAGATTTGAATGGAGCTGCAGAAAAGAAAACGAACGCTGGAATCCGAAG	4302
QY	4318	TGCAGTACACATCCAAATTAATGCAAAATCTGCCAAGTTGATTTTACTGTGACAACAATG	4377
Db	4303	TGCAGTATACATCTAACTAATGCAAAATCTGCCAAGTTGATTTTACTGTGACAACAATG	4362
QY	4378	GACTTTATACTGAGCCTCGCCCCATTTGGCAACCCGTTAACCCTGCCCTGTAAATTAC	4437
Db	4363	GACTTTATACTGAGCCTCGCCCCATTTGGCAACCCGTTAACCCTGCCCTGTAAATTGT	4422
QY	4438	GTGTTAATCAATAAACCGGTTGATTCGTTTCAGTTGAACCTTGCTCTGCTGCTTCTTA	4497
Db	4423	GTGTTAATCAATAAACCGGTTAATTGTTGTCAGTTGAACCTTGCTCTGCTGCTTATTA	4482
QY	4498	TCTTATC-GGTTACCATGCTTATAGCTTACACATTAATGCTTGTTGCGCTTCGAGTA	4556
Db	4483	TCTTATCTGTACACCAATAGCAACCGTTACACATTAATGCTTAGTTGCGCTTCGCGA--	4540
QY	4557	AAAGACTTACGTCATCGGGTTACCCCTAGTATGAGTTGCCCACTCCCTCTTCGCGCG	4616
Db	4541	-----ATACCCTAGTATGAGTTGCCCACTCCCTCTTAAGCGCGC	4581
QY	4617	TGCTGCTGCTGGTGGGCTGCGGACCAAAAGTCCGACAGACGGCAGAGCTCTGCTTGCC	4676
Db	4582	TGCTGCTGCTGGTGGGCGCGGACAGACAGAGCTCTGCGCTTGCGGACCTTGGTCCGCA	4641
QY	4677	GGCCCAACCGAGCGGACGACCGCGCAGAGAGGGAGTGGGCAA	4718
Db	4642	GGCCCAACCGAGCGGACGACCGCGCATAGAGGGAGTGGGCAA	4683

RESULT 3	
AAF23748	
ID	AAF23748 standard; DNA; 4722 BP.
XX	
AC	AAF23748;
XX	

DT 28-MAR-2001 (first entry)  
XX  
DE AAV3B DNA sequence.  
XX  
KM AAV3B; gene therapy; AAV viral vector; cystic fibrosis; cancer; AIDS;  
KM atherosclerosis; sickle cell anaemia; thalassaemia;  
KM blood clotting disorder; diabetes; ss.  
XX  
OS Adeno associated virus.  
XX  
PN US6156303-A.  
XX  
PD 05-DEC-2000.  
XX  
PF 11-JUN-1997; 97US-0873168.  
XX  
PR 11-JUN-1997; 97US-0873168.  
XX  
PA (UNIW ) UNIV WASHINGTON.  
XX  
PI Russell DW, Rutledge EA;  
XX  
DR WPI, 2001-060164/07.  
XX  
PT Adeno-associated virus serotype 6 and viral vector derived from it for  
PT gene therapy of cystic fibrosis, cancer, acquired immunodeficiency  
PT syndrome, sickle cell anemia, thalassemia and diabetes -  
XX  
PS Example 2; Fig 1; 50pp; English.  
XX  
CC The present invention relates to adeno-associated virus serotypes. The  
CC present sequence is the DNA sequence of one such serotype (AAV3B). AAV3B  
CC can be used to construct AAV viral vectors for use in gene therapy for a  
CC range of disorders: cystic fibrosis, cancer, AIDS, atherosclerosis,  
CC sickle cell anaemia, thalassaemia, blood clotting disorders and diabetes.  
CC The AAV viral vectors have increased transduction efficiency of a  
CC particular host cell as the AAV virion containing the AAV vector genome  
CC can be modified to express a capsid protein of an AAV serotype that  
CC transduces the selected host cell.  
SQ Sequence 4722BP; 1204 A; 1297 C; 1243 G; 978 T; 0 other;  
XX

Query Match	67.9%;	Score 3205;	DB 22;	Length 4722;
Best Local Similarity	81.6%;	Pred. No. 0;		
Matches 3872;	Conservative	0;	Mismatches 820;	Indels 55; Gaps 12;
QY 2	TGCCCACTCCCTCTCTGCGCGCTCGCTCGCTCGGTGGGGCCTGCGGACCAAGATCGCA	61		
DB 1	TGSCCACTCCCTCTATGCGCACTCGCTCGCTCGGTGGGGCTTGGCAGCAAAAGTCGCCA	60		
QY 62	GACGGCAGAGCTCTGCTCTGCGCGCCCAACCGAGCGAGCGCGCAGAGAGGAGTGG	121		
DB 61	GACGGACGTGCTTGACAGTCCGGCCCCACCGAGCGAGCGAGTGGCATAGAGGAGTGG	120		
QY 122	GCAACTCCATCACTAGGGGTATCGGAAGCGCTCCACGCTGCCGCTCAGCGCTGAC	181		
DB 121	CCAACCTCCATCACTAGAGGTATGCGAGTACGCTAACCGGAAGCGCGGGAAGCGAGACCAC	180		
QY 182	GTAATTACGTACATAGGGGAGTGTCTCTGTATTAAGCTGCACGTGAGTCTTTTGGACA	241		
DB 181	GCCTACCA-----GCTCGCTACGACGTACAGGTGA-CCCTTTTGGCACA	222		
QY 242	TTTTGGCACACCAAGTGGCCATTTAGGGTATATATGAGCCGAGTGACG-AGCAGGATCTC	300		
DB 223	GTTTGGCACACCAAGTGGCCGCTGAGGGTATATATTTCTGAGTGAGCGAAACGAGAGCTC	282		
QY 301	CATTTTGACCGCGAAATTTGAACGAGCAGCAGACCAATGCCGGGCTTCTACGAGATCGTGA	360		
DB 283	CATTTTGACCGCGAAATTTGAACGAGCAGCAGACCAATGCCGGGCTTCTACGAGATGTCCT	342		
QY 361	CAAGTGCCGAGCGAAGTGGACGAGCAGCACTGCCGGCATTTCTGACTCGTTTGAGCTG	420		
DB 343	GAAAGTCCCGAGTGAAGTGGACGAGCAGCAGCACTGCCGGCATTTCTAAGTCTGTTGTAAGT	402		



QY	421	GGTGGCCGAGAAAGAAATGGAGCTGCCCCCGGATTCTGACATGGATCTGAATCTGATTGA	480
Db	403	GGTGGCCGAGAAAGAAATGGAGCTGCCCGCGGATTCTGACATGGATCCGAATCTGATTGA	462
QY	481	GCAAGCACCCCCTGACCGTGGCCGAGAAAGCTGCAGCGCGACTTCTTGTTCCATGGGGCG	540
Db	463	GCAAGCACCCCCTGACCGTGGCCGAGAAAGCTTCAAGCGCGAGTCTTGTTGAGTGGCGCG	522
QY	541	CGTGAATAAGGCCCCCGAGGGCCCTCTTTTGTTCAGTTGAGAGGGCGAGTCTACTT	600
Db	523	CGTGAATAAGGCCCCCGAGGGCCCTCTTTTGTTCAGTTGAGAGGGCGAGTCTACTT	582
QY	601	CCACTCTCATATTCTGTGTGAGACCACGGGGGTCAAAATCCATGTGCTGGGCCGTTCT	660
Db	583	CCACTCTCACGTGCTGATTGAGACCATTGGGGGTCAAAATCCATGTGTGCTGGGCCGTTACT	642
QY	661	GAGTCAGATTAGGGAACAAGCTGTGTCAGAACCATCTACCGCGGATCGAGCCGACCTGCG	720
Db	643	GAGCCAGATTAAAGAGAGTGTGTACCCGATCTACCGCGGGTTCAGCCGCACTTCC	702
QY	721	CAACTGTTTCGGGTGACCAAGACGCGTAATGCGCCGAGCGGGGAACAAGTGTGTGA	780
Db	703	GAACGTGTTTCGGGTGACCAAAACGCGAATGCGCCGGGGTCGGGAACAAGTGTGTGA	762
QY	781	CGAGTGCTACATCCCACTACTCTCTGCCCAAGACTCAAGCTCCAGCTGCAGTGGGCGTG	840
Db	763	CGAGTGCTACATCCCACTACTCTCTGCCCAAGACTCAAGCTCCAGCTCCAGTGGGCGTG	822
QY	841	GACTTAACATGGAGGAGTATATAAGCGCTGTTGAACCTGGTCGAGCGCAACCGGCTGT	900
Db	823	GACTTAACATGGAGCAGTATTAAAGCGCTGTTGAATCTCGAGCGTAAACGGCTGTGT	882
QY	901	GCGCGAGCACTGACCCACGTCAGCGCAGACCCAGAGCAGACAAGAGAATCTGAACCC	960
Db	883	GCGCGAGCATCTGACGCACTGTTCGAGACGCAAGAGCAGACAAGAGAATCAGAACCC	942
QY	961	CAATTCTGACGCGCTGTCTATCCGGTCAAAAACTCCGGCGCTACATGAGCTGTGG	1020
Db	943	CAATTCTGACGCGCGGTCTATCAGTCAAAAACTCAGCCAGGTACATGAGCTGTGG	1002
QY	1021	GTTGGCTGTGGAACCGGGCATCACTCCGAGAGCAGTGGATCCAGAGACCAGGCTTC	1080
Db	1003	GTTGGCTGTGGAACCGGGGATCACTCAGAAAAAGCAATGGATTCAGAGACCAGGCTTC	1062
QY	1081	GTAACATCTCTTCAACGCGCTTCCAATCTCGCGGTCCAGATCAAGGCCGCTTGGACAA	1140
Db	1063	GTAACATCTCTTCAACGCGCTTCCAATCTCGCGGTCCAGATCAAGGCCGCTTGGACAA	1122
QY	1141	TGCCGGCAAGATCATGGCGCTGACCAAAATCCGCGCCGACTACTGTAGGCCCGCTCC	1200
Db	1123	TGCCCTCAAGATCATGAGCCTGACAAAGACGGCTCCGACTACTGTGTGGCAGCAACCC	1182
QY	1201	GCCCGGAGCATTTAAACCAACCGCATCTACCGCATCTGAGCTGAACGGCTACGAACC	1260
Db	1183	GCCCGAGAGCATTTACCAAAAATCGGATCTACCAATCTTGAGCTGAACGGGTACGATCC	1242
QY	1261	TGCCCTACCGCGGCTCCGCTTTCTTCGGCTGGGCCCAAGAAAGTTCCGGGAAGCGCAAC	1320
Db	1243	GCACTACCGCGGCTCCGCTTTCTTCGGCTGGGCCCAAAAGAGTTCCGGGAAGAGAACAC	1302
QY	1321	CATCTGGCTGTTTGGGCCGCAACAGGGCAAGACCAACATTCGGGAAGCCATGCGCCA	1380
Db	1303	CATCTGGCTGTTTGGGCCGCGCAAGCGGTAAACCAACATTCGGGAAGCCATGCGCCA	1362
QY	1381	CGCGGTGCCCTTCTACGCGCTGCGTCAACTGGAACCAATGAGACTTCCCTTCAATGATTG	1440
Db	1363	CGCGGTGCCCTTCTACGCGCTGCGTAAACTGGAACCAATGAGACTTCCCTTCAACGATTG	1422
QY	1441	CGTTCACAAGATGTTGATCTGTGTGGAGAGAGGCAAGATGACGCCCAAGTCTGTGAGTC	1500
Db	1423	CGTTCACAAGATGTTGATCTGTGTGGAGAGAGGCAAGATGACGCCCAAGTCTGTGAGAG	1482

QY	1501	CGCCAAAGGCCATTCTCGGCGGGCAGCAAGGTTGCGCGTGAACCAAAAGTGCAGTCTGCC	1560
Db	1483	CGCCAAAGGCCATTCTCGGCGGGCAGCAAGGTTGCGCGTGAACCAAAAGTGCAGTCTGCC	15422
QY	1561	CCAGATCGACCCCAACCCCGTGTATCGTCACTCCAACACCAACATGTGCGCGGTATGA	1620
Db	1543	CCAGATCGAACCCCACTCCCGTGTATCGTCACTCCAACACCAACATGTGCGCGGTATGA	1602
QY	1621	CGGGAACAGCACCACTTCGAGCACAGCAGCCGTTGAGAACCCGATGTTCAATTGA	1680
Db	1603	CGGGAACAGCACCACTTCGAGCATCAGCAGCCGCTGAGAACCCGATGTTAAATTGA	1662
QY	1681	ACTCACCCCGCGTCTGAGCATGACTTTGGCAAGTGCACAAAGCAGGAGTCAAGACTT	1740
Db	1663	ACTTACCCCGCGTCTGAGCATGACTTTGGCAAGTGCACAAAGCAGGAGTCAAGACTT	1722
QY	1741	CTTCGCGTGGGCGCAGATCACTGACCGAGGTGGCGCATGAGTTCTACGTCAAGAAAGG	1800
Db	1723	TTTCGCGTGGGCTTCGATCACTGACTGACGTGGCTCATGAGTTCTACGTCAAGAAAGG	1782
QY	1801	TGAGGCCAACAAGACCCGCCCGCATGACCGCGATTAAGCGAGCCCAAGCGGCGCTG	1860
Db	1783	TGAGACTAAGAAACGCCCGCTCCATGACCGCGATGTAAGCGAGCCCAAGCGGCGAGTG	1842
QY	1861	CCCCCTAGTCGCGGATCCATCGACGTACAGCGCGGAAAGAGCTCCGGTGACTTGC	1920
Db	1843	CACGTCACTTGCGCAGCGACACAACGTACAGCGCGGA--AGCACCGCGCGATACGCGGA	1899
QY	1921	CAGGTACCAAAACAATGTCTCTGTCACGCGGCGCATGCTTCAGATGCTGTTCCCTGCA	1980
Db	1900	CAGGTACCAAAACAATGTCTCTGTCACGCGGCGCATGATGCTTTCCTGTA	1959
QY	1981	GACATCGAGAGATGAATCAGAATTCACATTTGCTTCACGACGCGGACGAGAGACTG	2040
Db	1960	AACATCGAGAGATGAATCAAAATTTCCAATGTCTGTTTACGATGTTCAAGAGACTG	2019
QY	2041	TTCAGAGTGTCCCCCGCGGTGTCAGAATCTCAAC-----CGTGTGTCAAGAAAGGAC	2094
Db	2020	TGGGAATGTCTCCGTGAATGTCAGAATCTCAACCCGTTTCTGTGTCAAAAGAAAGAC	2079
QY	2095	GTAATCGAAACTCTGTGCCATTATCATCTGTGCGGCGGGCTCCGAGATTGCTGCTC	2154
Db	2080	TTATCAGAAACTGTGTCCAATTCATCATATCTGTGGAAGGGCACCCGAGATTGCTGTT	2139
QY	2155	GGCTGCGATCTGTTCACAGTGAACCTGATGACTGTGTTCTGAGCAATAAATGACTTA	2214
Db	2140	GGCTGCGATTTGGCCAATGTGACTTGATGACTGTGTTCTGAGCAATAAATGACTTA	2199
QY	2215	AACAGGTATGGCTGCCATGTGTTATCTTCCAGATTGGCTCGAGACAACCTCTGAGG	2274
Db	2200	AACAGGTATGGCTGTCTGACGCTTATCTTCCAGATTGGCTCGAGACAACCTTCTGAAG	2259
QY	2275	GCAATTCGCGAGTGTGTGGAATTGAAACCTTGAGACCCCGAAAGCCCAAGCCACGCAAA	2334
Db	2260	GCAATTCGTGAGTGTGTGGAATTGAAACCTTGAGTCCCTCAACCCAAAGCGAACCAACAAC	2319
QY	2335	AGCAGGACGACGCGCGGGGTCTGTGCTTCTGTGCTACAAGTACCTCGGACCCCTTCAACG	2394
Db	2320	ACCAGGACAACCGTCGGGCTCTGTGCTTCCGGTTACAATACTCGGACCCGTTAACG	2379
QY	2395	GACTCGACAAGGGGGAGCCGTCACACGCGCGGACGACGCGCCCTCGAGCAGCAAGG	2454
Db	2380	GACTCGACAAGGAGAGCCGTCACACGAGCGGACGCGGACCCCTCGAACACGACAAAG	2439
QY	2455	CCTACGACGACGACTCAAGCGGGTGACAATCCGTAACCTGCGGTATAACCAAGCGGACG	2514
Db	2440	CTTACGACGACGACTCAAGCGGGTGACAACCCGTAACCTCAAGTACAACCAAGCGGACG	2499
QY	2515	CCGAGTTTCAAGAGCGTCTGCAAGAAGATACGTCTTTTGGGGGCAACCTCGGCGGACGAG	2574
Db	2500	CCGAGTTTCAAGAGCGTCTTCAAGAAGATACGTCTTTTGGGGGCAACCTTGGCAGAGCAG	2559
QY	2575	TCTTCCAGGCCAAGACGGGTTCTCGAACCTCTCGGTCTGTTGAGGAAGCGGCTAAGA	2634

Db 2560 TCTTCCAGGCCAAAAGAGGATCTTGAGCCTCTGTGTGTGAGGAGACGCTAAAA 2619  
Qy 2635 CGGCTCCTGGAAGAAGACGTCCGGTAGACATCGCCACAAGAGCCAGACTCCTCTCGG 2694  
Db 2620 CGGCTCCTGGAAGAAGAGCGCTGTAGATCACTCTCAGGAACCCGAGCTCATCATCTG 2679  
Qy 2695 GCATCGGCACAGACAGGCCAGCAGCCCCGCTAAAAAGAGACTCAATTTTGGTCAAGTGGCG 2754  
Db 2680 GTGTGGCAAAATCGGGCAAAACAGCCTCGCCAGAAAAAGACTAAATTTGGTCAAGTGGCG 2739  
Qy 2755 ACTCAGAGTCAGTCCCGATCCACAACCTCTCGGAGAACCTCCAGCAACCCCGCTGTG 2814  
Db 2740 ACTCAGAGTCAGTCCCGATCCACAACCTCTCGGAGAACCTCCAGCAACCCCGCAAGTT 2799  
Qy 2815 TGGGACCTACTACAATGCTTCAAGCGGTGCGCACCAATGCGACAGCAATAACGAGGCG 2874  
Db 2800 TGGGATCTAATAACAATGCTTCAAGCGGTGCGCACCAATGCGACAGCAATAACGAGGCG 2859  
Qy 2875 CCGACGAGTGGGTATGCTCCTCAGAAATGGCATTTGCATTCAGATTCAGTGGCGGACA 2934  
Db 2860 CCGATGGAGTGGGTATTTCTCAGGAATTTGGCATTTGCATTCAGATTCAGTGGCGGACA 2919  
Qy 2935 GAGTCATACCAACCAAGCAGCAGCCGCACTTGGGCTTGGCCACTTACATAACCACTTACA 2994  
Db 2920 GAGTCATACCAACCAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2979  
Qy 2995 AGCAATCTCCAGTGTCTTCAACGGGGGCGACAAAGCAACCACTACTTGGCTACAGCA 3054  
Db 2980 AGCAATCTCCAGTGTCTTCAACGGGGGCGACAAAGCAACCACTACTTGGCTACAGCA 3036  
Qy 3055 CCCCCTGGGGTATTTGATTTCAACAGATTCACCTGCCACTTTTCAACCACTGACTGGC 3114  
Db 3037 CCCCCTGGGGTATTTGATTTCAACAGATTCACCTGCCACTTTTCAACCACTGACTGGC 3096  
Qy 3115 AGGACTCATCAACAATTTGGGATTCGGGCCCAAGAGACTCAACTTCAAACTTTCA 3174  
Db 3097 AGGACTCATCAACAATTTGGGATTCGGGCCCAAGAGACTCAACTTCAAACTTTCA 3156  
Qy 3175 ACATCCAGTCAAGAGGTCAAGCAAGATGATGGCGTCAACCACTGCTAATAACCTTA 3234  
Db 3157 ACATCCAGTCAAGAGGTCAAGCAAGATGATGGCGTCAACCACTGCTAATAACCTTA 3216  
Qy 3235 CCAGCAGGTTCAAGTCTTCTGGACTCGGAGTACCAGCTTCCGTACGTCTCGGCTGTG 3294  
Db 3217 CCAGCAGGTTCAAGTCTTCTGGACTCGGAGTACCAGCTTCCGTACGTCTCGGCTGTG 3276  
Qy 3295 CGCACCAGGCTGCTCCTCCCTCCGTTCCCGGCGGAGCGTTCATGATTCGCAATACGGCT 3354  
Db 3277 CGCACCAGGCTGCTCCTCCCTCCGTTCCCGGCGGAGCGTTCATGATTCGCAATACGGCT 3336  
Qy 3355 ACCTGACGCTCAACATGCGAGCCAGCCGAGGAGCGTTCACTCTTTACTGCTGGAAT 3414  
Db 3337 ACCTGACGCTCAACATGCGAGCCAGCCGAGGAGCGTTCACTCTTTACTGCTGGAAT 3396  
Qy 3415 ATTTCCCTTCTCAGATGCTGAGAACGGGCAACAACCTTACCTTCACTACACTTTGAGG 3474  
Db 3397 ACTTCCCTTCTCAGATGCTGAGAACGGGCAACAACCTTCACTACACTTTGAGG 3456  
Qy 3475 AAGTGCCTTTCCACAGCAGCTACCGCGCACGCGAGGCTGAGCGGCTGATGATCTC 3534  
Db 3457 AATGACCTTTTCCACAGCAGCTACCGCGCACGCGAGGCTGAGCGGCTGATGATCTC 3516  
Qy 3535 TCATGACCAATACCTGTATTAACCTGAAGAACTCA--AAATCAGTCCGGAAGTGGCC 3591  
Db 3517 TTATGATCAATATCTGTACTACTGAACAGAAAGCAACCACTCTGGAACCAACA 3576  
Qy 3592 AAAACAAGGACTGCTGTTAGCGGTGCTCCAGCTGCGATGCTGTTCAAGCCCAAA 3651  
Db 3577 ACCAATCAGGCTGCTTTTAAAGCAGGCTGGGCTCAGTCTATGCTTTGACAGCCAGAA 3636  
Qy 3652 ACTGGCTACTGAGCCCTGTTATCGGACAGCGCGTTTCAAAAACAAGACAGACACA 3711

Db 3637 ATTGGCTACTGGGCCCTGCTTACCGGCAACAGAGACTTTCAAGACTGCTAACGACACACA 3696  
Qy 3712 ACAACAGCAATTTTACCCTGAGCTGCTGCTTCAAAATATTAACCTCAATGGCGGTATCCA 3771  
Db 3697 ACAACAGTAACTTCTTGGAGAGCGGCCAGCAAAATATCACTCAATGGCGCGACTCGC 3756  
Qy 3772 TCATCAACCTGGCACTGCTATGGCTCAACAAAGACGACGAGCAAGTCTTTCCCA 3831  
Db 3757 TGGTGAATCCAGGACGAGCTATGGCCAGTCAAGAGCGATGAAGAAAAATTTTCCCTA 3816  
Qy 3832 TGAGCGGTCTATGATTTTGGAAAAAGAGAGCGCCGAGCTTCAACAACCTGCATTGGACA 3891  
Db 3817 TGCACGGCAATCTAATATTTGGCAAGAGGACCAACGGCAAGTAAACGCAATTTAGATA 3876  
Qy 3892 ATGTCAATGATTAACAGCAGAGGAAATTAAGCCACTAACCTGTGGCCACCGAAAGAT 3951  
Db 3877 ATGTAAATGATTAACGATGAAGAGAGATTCGTACCAACCAATCTGTGGCAACAGAGAGT 3936  
Qy 3952 TTGGACCGTGGCAGTCAATTTCCAGACGACGACGACGACGACGACGACGACGAGATGTC 4011  
Db 3937 ATGGAATGTGCAAAATTAACCTGAGAGCTCAAAATACAGCTCCACAGACTAGACTGTCA 3996  
Qy 4012 ATGCTATGGAGCATTAACCTGGCATGGTGTGGCAAGATAGAGCGTGTACTGACGGTTC 4071  
Db 3997 ATGATCAGGGGGCTTACCTGGCATGGTGTGGCAAGATGCTGTACTTCAAGAGAC 4056  
Qy 4072 CCATTTGGGCCAAATTCCTCAACAGATGAGACACTTCAACCCGCTCTCTTATGGCG 4131  
Db 4057 CTATCTGGGCAAAAGATTCCTCAACGAGATGAGACACTTCACTCTCTCTGATGGGAG 4116  
Qy 4132 GCTTTGACTCAAGAACCCGCTCTCAGATCCTCATCAAAAACAGCCTGTCTCTGCGA 4191  
Db 4117 GCTTTGACTCAAGAACCTCGCTCTCAATCATGATCAAAATACTCCGCTACCGGCA 4176  
Qy 4192 ATCTCCGGCGAGTTTCACTCAAAAGTTTGGCTTCACTCATCAACCAATACCTCCACAG 4251  
Db 4177 ATCTCCGACGACTTTCAGCCCGGCAAGTTTGGCTTCACTCATCACTCAGTACTCCACTG 4236  
Qy 4252 GACAGTGAAGTGAATTTGAATGGAGCTGCAGAAAGAAACAGCAAGCGCTGGAATC 4311  
Db 4237 GACAGTGAAGTGAATTTGAATGGAGCTGCAGAAAGAAACAGCAAGCGTGAATC 4296  
Qy 4312 CCGAATGCAGTACACATCCAAATTAAGCAAAATCTGCCAAGTGAATTTACTGTGACA 4371  
Db 4297 CAGAGATTCAGTACACTTCCAACTAACAAAGTCTGTAAATGTGACTTACTGTAGACA 4356  
Qy 4372 ACAATGCACTTATCTAGCTCGCCCATTTGGCACCCTGTTACCTTACCCTCCCTGT 4431  
Db 4357 CTAAATGCTTTATAGTGAACCTCGCCCTAATTTGAACCCGCTATCTCACACGAACTTGT 4416  
Qy 4432 AATTAAGTGTAAATCAATTAACCGGTTGATTCGTTTCACTTGAACCTTTGG-TCCTCTGTC 4490  
Db 4417 AA-TCCTGTTAATCAATTAACCGTTTAATTCGTTTCACTTGAACCTTTGGCTTGTGCA 4475  
Qy 4491 CTCTTATCTTATC-GTTTACAGTGTATAGCTTACATTAATCTG-----CTTGGTT 4543  
Db 4476 CTCTTATCTTATCTTGTTCAGATGGCTACTGCGTAGATTAAGCAGCGGCTGCGGCGCTT 4535  
Qy 4544 GCGCTTCGCGATAAAGAC-----TTACGTCATCGGGTTAACCCCTAGTATG 4591  
Db 4536 GCGCTTCGCGGTTTACAACTGCTGTATTAATTTTAACTCTCGCCATACTCTAGTATG 4595  
Qy 4592 AGTTGCCCACTCCCTCTCTGCGGCTCGCTGCTCGGTGGGGCTGCGGACCAAGGTTC 4651  
Db 4596 AGTTGCCCACTCCCTCTCTAGGCACTCGCTCGCTCGGTGGGGCTGCGGACCAAGCAAG 4655  
Qy 4652 GCAGACGGCAGAGCTGCTGCTGCGGCGGCCCAACCGAGCGAGCGGACGAGAGGAG 4711  
Db 4656 TCCGTGTGGCAGCTTGTGTGCGCAGGCGCCACCGAGCGAGCGAGTGCATAGAGGAG 4715  
Qy 4712 TGGGCAA 4718  
Db 4716 TGGCCAA 4722





OY		717	TGCCCAACTGTTTCGGGTGAACCAAGACCGCTAATGCGCCGAGGGGGAAACAAGTGG	776
Db		703	 TGCCAAACTGTTTCGGGTGAACCAAGACCGCTAATGCGCCGAGGGGGAAACAAGTGG	762
OY		777	TGACAGTAGTCTACATCCCCCACTAACCTCCTGCCCCAAGACTCAGCCCAGCTGCAGTGG	836
Db		763	 TGATAGTAGTCTACATCCCCCACTAACCTCCTGCCCCAAGACTCAGCTCCAGTGG	822
OY		837	CGTGAATAACATGAGGAGTATTATAAGCCCTGTTGAACCTGGCCGAGCCGAACCGC	896
Db		823	 CGTGAATAATATGGAACAGTATTTTAAGCCCTGTTGAATCTCACGGAGCCTAAACGCT	882
OY		897	TGCTGGCGCAGCACCTGACCACGTCAAGCCAGACCCAGAGACGAACAAGAGAATCTGA	956
Db		883	 TGCTGGCGCAGCATCTGACGCACTGTCTGACAGACGACGAGACGAACAAGAGAATCAGA	942
OY		957	AACCCAATTCTGACGCGCCTGTATCCGGTCAAAAACTCCGCGCTACATGAGCTGG	1016
Db		943	 ATCCCAATTCTGATGCGCGGTGATCAGATCAAAAACTTCAGCCAGGTACATGAGCTGG	1002
OY		1017	TCCGGTGGCTGTGACCCGGGGCATCACTCCGAGAAGCAGTGATCCAGAGACCAAG	1076
Db		1003	 TCCGGTGGCTGTGACCAAGGGGATTAACCTCGAGAAGCAGTGATCCAGAGACCAAG	1062
OY		1077	CCTCGTACATCTCCTTCAACGCCCTTCCAACTCGCGGTCCAGATCAAGGCCCTCTGG	1136
Db		1063	 CCTCATACATCTCCTTCAATGCGGCTCCAACTCGCGGTCCAAATCAAGGCTGCTTGG	1122
OY		1137	ACAATGCCGCAAGATCATGGCGCTGACCAAAATCCGCGCACTACTGTGAGCCCG	1196
Db		1123	 ACAATGCCGGAAGATTATGAGCTGACTAAAACCGCCGACTACTGTGGGCCAGC	1182
OY		1197	CTCCGCCGCGACATTAACCAACCGCATCTACCGCACTTGAGCTGAACGGCTACG	1256
Db		1183	 AGCCCGTGAGGACATTTCCAGCAATCGGATTTATAAAATTTTGAACTAAACGGGTACG	1242
OY		1257	AACCTGCCTACGCCGCTCCGCTTTCTCGGCTGGGCCAGAAAAGTTCCGGAAGCGCA	1316
Db		1243	 ATCCCCAATATGCGGCTTCCGCTTTCTGCGATGGGCCACGAAAAAGTTCCGCAAGAGCA	1302
OY		1317	ACAACCATCTGGCTGTTTGGGCGGCCCAACGCGGCAAGACCAATCGCGGAAGCCATCG	1376
Db		1303	 ACAACCATCTGGCTGTTTGGGCTGCAACTACCGGAAGACCAATCGCGGAGGCCATAG	1362
OY		1377	CCCAACGCGTGCCCTTCTACGGCTGCGTCAACTGACCAATGAGAACTTTCCCTCAATG	1436
Db		1363	 CCCAACTGTGCCCCCTTCTACGGGTGCGTAAACTGACCAATGAGAACTTTCCCTCAACG	1422
OY		1437	ATTGCGTCGACAAGATGATGATCTGTTGGAGAGGGCAAGATGACCGCCAAAGTCTGG	1496
Db		1423	 ACTGTGTGACAAGATGATGATCTGTTGGAGAGGGGAAGATGACCGCCAAAGTCTGG	1482
OY		1497	AGTCCGCCAAGGCCATTCTCGCGCGGACGCAAGGTGCGCGTGAACCAAAAGTGCAAGTCT	1556
Db		1483	 AGTCCGCCAAGGCCATTCTCGSAGGAAGCAAGGTGCGCGTGAACCAAAATGCAAGTCT	1542
OY		1557	CCGCCACGATGACCCCAACCCCGTGATGTCACCTCCAACCAACATGTGCGCCGTGA	1616
Db		1543	 CGGCCACGATGACCCCACTCCGTGATGTCACCTCCAACCAACATGTGCGCCGTGA	1602
OY		1617	TGACGGGAACAGCACCACTTCGAGCACCAAGCAGCCGTTGACGACCGGATGTTCAAAT	1676
Db		1603	 TGACGGGAACACTCAACGACCTTCGAACACCAAGCAGCCGTTGCAAGCCGATGTTCAAAT	1662
OY		1677	TTGAACTCACCCGCGTCTGAGCATGACTTTGSCAAGGTGACAAAGCAGGAAGTCAAAG	1736
Db		1663	 TTGAACTCACCCGCGTCTGATCATGACTTTGSGAAGGTGACCAAGCAGGAAGTCAAAG	1722
OY		1737	AGTTCTTCCGCTGGGGCGAGATCACTGACCGAGGTGGCGCATGAGTTCTACGTACAA	1796
Db		1723	 ACTTTTCCGCTGGGCAAGGATCACTGTGAGGTGAGCATGAATTTACGTACAAA	1782

QY	1797	AGGGTGGAGCCCAACAAAAGACC	CGCCCCCGATGACCGGGATAAAGCGAGCCCAACGGG	1856
Db	1783	AGGGTGGAGCCCAAGAAAAGACC	CGCCCCCAGTGAAGCATATATAGTAGCCCAACGGG	1842
QY	1857	CCTGCCCCCTAGTCGCGGATCCAT	CGAGCTCAGACCGGAAGAGCTCCGGTGACTTG	1916
Db	1843	TGGCGAGTCACTTGGCGAGCCAT	CGAGCTCAGACCGGAAGAGCTCCGGTGACTTG	1899
QY	1917	CCGACAGGTACCAAAACAATGTT	CTCGTACGCGGGCATGCTTCAGATGCTGTTCCCT	1976
Db	1900	CAGACAGGTACCAAAACAATGTT	CTCGTACGCGGGCATGGAATCTGATGCTGTTCCCT	1959
QY	1977	GCAAGACATGCGAGAGATGAAT	CAGAAATTCAAATTTGCTTCACGACGGGAGAGAG	2036
Db	1960	GCAGACAATGCGAGAGATGAAT	CAGAAATTCAAATTTGCTTCACGACGGGAGAGAG	2019
QY	2037	ACTGTTCAAGTGTCTCCCGCGT	GTGCAAAATTCACCCGGT--CGTCAGAAAGAGA	2093
Db	2020	ACTGTTTAAAGTGTCTTCC--	CGTGTCAAGATTCACCCGGTCTGTCGTCAAAAAGG	2076
QY	2094	CGTATCGGAACTCTGTGCCAT	TCATCATCTGTGCGGCGGGCTCCCGAGATTGCTT	2153
Db	2077	CGTATCAGAACTGTGTCTACAT	TATCATATCATGGGAAAGGTGCCAGA--CGTTTGA	2133
QY	2154	CGGCTCGCATCTGTCAACGT	GACCTGGATGATCTGTCTTGAGCAATAATGACTT	2213
Db	2134	CTGCTCGCATCTGTCAATGT	GATTTGGATGATCTCATCTTTGAAACAATAATGATTT	2193
QY	2214	AAACGAGTATGGCTGCCGAT	GTATCTTCCAAATTTGGCTCGAGACAACCTCTCTGAG	2273
Db	2194	AAATCAGTATGGCTGCCGAT	GTATCTTCCAAATTTGGCTCGAGACACTCTCTCTGA	2253
QY	2274	GGCATTCGCGAGTGTGGGA	CTTGAACCTTGAGCCCGGAAGCCCAAGCCCAACGACA	2333
Db	2254	GGAATTAAGACAGTGTGGA	AGCTCAAACTTGCGCCACCAACCAAGCCCGCAGACGG	2313
QY	2334	AAGCAGGACGACGGCGGGGT	CTGTGCTTCTGCTACAAGTACCTCGACCTTCAAC	2393
Db	2314	CATAAGACGACGACGGGCT	TGTGCTTCTGCTACAAGTACCTCGACCTTCAAC	2373
QY	2394	GGACTCGACAAGGGGGAG	CCCTCAACGCGCGGAGCAGCGGCCCTCGAGCAGCAAG	2453
Db	2374	GGACTCGACAAGGGAGAC	CGGTCAACGAGGCAAGCGCGGCCCTCGAGCAGCAAA	2433
QY	2454	GCCTACGACCGACAGCT	CAAAGCGGGTGACAATCCGTAACCTCGGTATAACACCGCCAG	2513
Db	2434	GCCTACGACCGGCACT	CGAAGCGGAGACAACCCGTAACCTCAAGTACAACACCGCCAG	2493
QY	2514	GCCGAGTTTCAAGAGCG	CTCTGCAAGAAGATACGCTTTTGGGGGCAACCTCGGGCGAGA	2573
Db	2494	GCCGAGTTTCAAGAGCG	CTTAAAGAAGATACGCTTTTGGGGGCAACCTCGAGCAGCA	2553
QY	2574	GTCCTTCCAGGCGCAAGA	AGCGGTTCTCGAACCTCTCGGTCTGGTTAGGAAGCGCTAAG	2633
Db	2554	GTCCTTCCAGGCGA	AAAAGAGGGTTCTTGAAACCTCTGGGCTCTGTTAGGAACCTGTTAAG	2613
QY	2634	ACGGCTCCTGGAAGAA	AGCTCCGGTAGAGCAGTCCGCAAGAGCCAGACTCTCTCTCG	2693
Db	2614	ACGGCTCCGGGA	AAAAAGAGGCCGTTAGAGCACTCTCTGTGAGCCAGACTCTCTCTCG	2673
QY	2694	GCGATCGGCAAGACAG	CGCCAGCGCTAAAAAGAACTCAATTTTGGTCAAGTGGC	2753
Db	2674	GGAACCGGAAGCGGG	CCAGCAGCTGCAAGAAAAAGATTGAATTTTGTCAAGCTGGA	2733
QY	2754	GACTCAGAGTCACT	CCCCGATCCACAACCTCTCGAGAAACCTTCAGCAACCCCGCTGCT	2813
Db	2734	GACGACAGACTCAG	TACCTGACCCCGACCTCTCGGACAGCCACGACGCCCTCTGCT	2793
QY	2814	GTGGACCTACTACA	TATGGCTTCAGGCGGTGGCGCAACCAATGGCAGACAATAAGAGGC	2873
Db	2794	CTGGGA	ACTAATACGATGGCTTACAGGCAAGTGGCGCAACCAATGGCAGACAATAAGAGGC	2853
QY	2874	GCCGACGAGTGGG	TATGCCCTCAGGAATTTGGCAATTCGATTCACATGCTGGGGCGAC	2933

```
|||||
Db 2854 GCCGACGGAGTGGTAATTCCTCGGAAATTTGGCATTCGATTCACATGGATGGCGAC 2913
Qy 2934 AGAGTCATCACACCAGACACCCGACCTGGGCTTGCCACCCTACAATAACCACTCTAC 2993
Db 2914 AGAGTCATCACACCAGACACCCGAACTGGGCTTGCCACCCTACAATAACCACTCTAC 2973
Qy 2994 AAGCAATCTCCAGTGTTCAAACGGGGCCAGCAAGCAACCACTACTTGGCTACAGC 3053
Db 2974 AAACAAATTTCCAGCCAATCA--GAGAGCTCGAAGCAATCACTACTTGGCTACAGC 3030
Qy 3054 ACCCGCTGGGTAATTTGATTTCAACAGATTCCACTGCCACTTTTACCAAGTGTGG 3113
Db 3031 ACCCGTGGGGTAATTTGACTTCAACAGATTCCACTGCCACTTTTACCAAGTGTGG 3090
Qy 3114 CAGGACTCATCAACAATTTGGGATTCGGGCCAAGAGACTCAACTTCAACTCTTC 3173
Db 3091 CAAAGACTCATCAACAACACTGGGATTCGCAACCAAGAGACTCAACTTCAAGCTCTT 3150
Qy 3174 AACATCCAAGTCAAGAGGTCACGACGAATGATGGCTCAACAACCATCGCTAATACTT 3233
Db 3151 AACATTCAAAGTCAAGAGGTCACGACGAATGATGGCTCAACAACCATCGCTAATACTT 3210
Qy 3234 ACCGACCGGTTCAAGTCTTCTCGACTCGGAGTACCAAGCTTCCGTACTCGGCTCT 3293
Db 3211 ACCGACCGGTTCAAGTCTTCTCGACTCGGAGTACCAAGCTTCCGTACTCGGCTCG 3270
Qy 3294 GCGCACCAGGCTGCTCCCTCCCTCCGTCGCGGAGCTGTTCATGATTCGCAATACGCG 3353
Db 3271 GCGCATCAAGGATGCTCCCGCCGTTCCAGCAGAGCTTTCATGTGTGCCACAGTATGA 3330
Qy 3354 TACCTGACGCTCAACAATGGCAGACCAAGCCGTGGAGCTTCATCTTTTACTGCTGAA 3413
Db 3331 TACCTACCGCTGAACAAGGAGTCAAGCAGTAGAGAGCTTTCATTTTACTGCTGAG 3390
Qy 3414 TATTTCCCTTCTAGATGCTGAGAACGGGCAACAATTACCTTCAGCTACACTTTGAG 3473
Db 3391 TACTTCTCTCAGATGCTGCGTACCGGAACAATTACCTTCAGCTACACTTTGAG 3450
Qy 3474 GAAATGCTTTCCACAGACGCTACGGGCAACAGCCAGAGCTGAGCCGGTGAATGCT 3533
Db 3451 GAAATGCTTTCCACAGACGCTACGGGCAACAGCCAGAGCTGAGCCGGTGAATGCT 3510
Qy 3534 CTCATGACCAATACCTGTATTACCTGAACAGAACTCAAAATCAGTCCGGAAGTCCCAA 3593
Db 3511 CTCATGACCAATACCTGTATTACCTGAACAGAACTCAAAATCAGTCCGGAAGTCCCAA 3570
Qy 3594 AACAAAGACTTGTGTTTAAAGCCGTGGGTCTCCAGCTGGCATGTCTGTTCAAGCCCAA 3653
Db 3571 CAGTCAAGGCTTCAAGTTTCTCAGGCGGAGCGAGTGAATTCGGGACGATCTTGAAG 3630
Qy 3654 TGGTACCTGACACCTGTATCGGACAGCGGCTTTCTAAACAAAAACAGACAAC 3713
Db 3631 TGGTTCCTGACACCTGTATCGGACAGCGGATATCAAGACATCTGGGATTAACAAC 3690
Qy 3714 AACAGCAATTTTACTGGAAGTGTGCTTCAAAATATACTCAATGGGCGTGAATCCATC 3773
Db 3691 AACAGTGAATATCTGTGAGACTGAGCTACCAAGTACCACTCAATGGGAGAGACTCTCTG 3750
Qy 3774 ATCAACCTGGAAGTGTATGAGCTTCAACAAGAGAGAGAGAGAGAGAGAGAGAGAG 3833
Db 3751 GTGAATCCGGGCGGCGCATGGCAAGCCACAAGAGAGAGAGAGAGAGAGAGAGAGAG 3810
Qy 3834 AGCGGTGTATGATTTTGGAAAAAGAGAGCGCGGAGCTTCAACACTGATGAGACAAT 3893
Db 3811 AGCGGGTTCATCTTTGGGAAGCAAGGCTCAAGAAAAAATAATGTGCAATTGAAAG 3870
Qy 3894 GTCATGATTACAGAGAGAGAAATTAAGCCACTAACCTGTGCGCAACGAAAGATTT 3953
Db 3871 GTCATGATTACAGAGAGAGAAATCAAGCAACCAATCCGTGCTACGAGAGAGATAT 3930
Qy 3954 GGGACCGTGCAGTCAATTTCCAGAGACAGACAGAACCCCTGCGACCGAGATGTGCAT 4013
|||||
```

```
Db 3931 GGTTCGTATCTACCAACCTCCAGAGAGGCAACAGACAAGAGACTACCGAGATGTCAAC 3990
Qy 4014 GCTATGGAGCATTACCTGGCATGCTGTGGCAAGATAGAGAGTGTACTGACAGGTCCC 4073
Db 3991 ACACAAGGGTCTTCCAGGCAATGTCTGGCAGAGACAGAGATGTACTTACAGGGGCC 4050
Qy 4074 ATTTGGCCAAATTCCTCAACAGATGAGACACTTTCACCGCTCTCTTATGGGCGGC 4133
Db 4051 ATCTGGGCAAGATTCCACACAGGACGAGACATTTTCAACCCCTCTCCCTCATGGGTGA 4110
Qy 4134 TTTGACTCAAGAACCCGCTCTCAGATCTCTCATCAAAAACAGCCCTGTCTCTCGCAAT 4193
Db 4111 TTGGAATTAACACACCTCTCCAGAGATTTCTCATCAAGAACACCCCGTACTGCGAAT 4170
Qy 4194 CCTCCGGGAGTTTTCAGCTACAAAGTTTGTCTTCATTCATCAACCAATACTCCACAGA 4253
Db 4171 CCTTCGACCACTTCAGTGGCGCAAGTTTGTCTTCTTCATCAACAGTACTCCACGGGA 4230
Qy 4254 CAAGTGAAGTGAATAATGAATGGAGCTGCAGAAAGAAAACAGCAAGCGCTGGAATCCC 4313
Db 4231 CAGTCAAGCTGAGATCGAGTGGAGCTGCAGAAAGAAAACAGCAAGCGCTGGAATCCC 4290
Qy 4314 GAAGTGAAGTACATCCAAATTAAGCAAAATCTGCCAAGTGAATTTACTGTGACAAAC 4373
Db 4291 GAAATTCAGTACACTTCCAACTACAAACAGTCTGTTAATGTGACTTTACTGTGACACT 4350
Qy 4374 AATGACTTTATCTAGAGCTCGGCCATTTGGCACCCGTTACTTACCCTGTAA 4433
Db 4351 AATGGCGTATTCAGAGCTCGGCCATTTGGCACAGATACCTGACTGTAATCTGTAA 4410
Qy 4434 TTACGTGTTAATCAATAACCGGTTGATTCGTTTCAAGTTGAATTTGCTCTCTCTT 4493
Db 4411 TTGCTGTTAATCAATAACCGTTAATTCGTTTCAAGTTGAATTTGCTCTCTCTT 4470
Qy 4494 CTATCTTATCGGTTTACCATGTTATAGCTTACACATTAATGCTTGTGCTGCTGCG 4553
Db 4471 TCTTCTTATCTAGTTTCCATGCTTACGATTAAGTATGAGTGGCGGTTAATCATTAAC 4530
Qy 4554 ATAAAGACTTACGTCATCGGTTTACCCCTAGTATGATGAGTTGCCACTCCCTCTGCG 4613
Db 4531 TACAAGG-----ACCCCTAGTATGATGAGTTGCCACTCCCTCTGCG 4574
Qy 4614 CGCTGCTGCTCGGTGGGCGCTGCGGACCAAGGTCCGACAGCGGAGAGCTGTGCTCT 4673
Db 4575 CGCTGCTGCTCACTGAGGCGGCGGACCAAGGTCCGCGGAGCGCTGCTGCTGCG 4634
Qy 4674 GCCGGCCCAACCGAGCGAGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4718
Db 4635 GCGGCGCTCAGTGAAGCGAGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4679
|||||

RESULT 5
AAH26325
ID AAH26325 standard; DNA; 8698 BP.
XX
AC AAH26325;
XX
DT 02-OCT-2001 (first entry)
XX
DE Wild-type adeno-associated virus genome plasmid pAV2.
XX
KW AAV; vector; pAV2; packaging cell line; gene therapy; ds.
XX
OS Adeno associated virus.
XX
PN WO200155361-A2.
XX
PD 02-AUG-2001.
XX
PF 26-JAN-2001; 2001WO-US02709.
XX
PR 26-JAN-2000; 2000US-0178536.
XX
```

PA (CHIR ) CHIRON CORP.  
 XX  
 PI Hardy SF;  
 XX  
 DR WPI; 2001-483239/52.  
 XX  
 PT Producing recombinant adeno-associated virus (rAAV) vector, by stably  
 PT infecting eukaryotic host cell with rAAV vectors, helper adenovirus or  
 PT herpes virus, recombinant adenovirus or herpes vectors -  
 XX  
 PS Example 1; Page 51-54; 63pp; English.  
 XX  
 CC The present sequence is that of adeno-associated virus (AAV)  
 CC packaging plasmid PAV2, which contains a wild-type AAV type 2  
 CC genome. The following AAV packaging plasmids were prepared:  
 CC pKsrepCap, a non-replicating control plasmid containing rep and  
 CC cap genes but no inverted terminal repeats (ITRs); PAV2; and  
 CC pFloxAAV (see AAH26326), a version of PAV2 in which 2 inserted loxp  
 CC sites flank the rep and cap genes and thus separate the ITRs from  
 CC these genes. Each was combined with pCMV GFP (see AAH26324), a  
 CC plasmid with a green fluorescent protein expressing recombinant AAV  
 CC (rAAV) genome, and used to transfect 293 or Cre8 cells. After 6  
 CC hr, the medium was changed and wild-type adenovirus type 5 was  
 CC added. Virus particles were harvested 3 days later. Plasmid PAV2  
 CC produced mostly AAV and a low yield of GFP vector with no effect  
 CC of Cre recombinase. pFloxAAV packaged as much GFP vector as the  
 CC control plasmid and while the amount of GFP vector was not  
 CC affected by Cre recombinase, the amount of floxAAV in the product  
 CC was reduced to 1/10 of the 293 value by the action of Cre  
 CC recombinase. In 293 cells it was evident that the GFP vector did  
 CC not compete with wild-type AAV for replication or packaging, while  
 CC the GFP vector was able to complete against floxAAV. Methods  
 CC and compositions are provided for producing rAAV vector particles  
 CC by: (a) introducing into a host cell (i) pFloxAAV, (ii) a  
 CC recombinant viral vector encoding plasmid, and (iii) a plasmid  
 CC encoding herpesvirus, cytomegalovirus or adenoviral functions, or  
 CC a herpesvirus, cytomegalovirus or adenovirus itself, to produce  
 CC flox AAV particles and rAAV particles; and (b) introducing into a  
 CC second host cell (i) the rAAV particles or (a), (ii) a vector  
 CC that directs expression of Cre, and (iii) a vector which directs  
 CC expression of herpes virus, cytomegalovirus or adenovirus helper  
 CC functions, such that rAAV vector particles are produced. The  
 CC vectors are useful for in vivo or in vitro gene therapy and also  
 CC for in vitro recombinant protein production.  
 XX  
 SQ Sequence 8698 BP; 2117 A; 2368 C; 2306 G; 1907 T; 0 other;  
 Query Match 65.9%; Score 3107.8; DB 22; Length 8698;  
 Best Local Similarity 80.2%; Pred. No. 0;  
 Matches 3790; Conservative 0; Mismatches 882; Indels 53; Gaps 10;

Db 283 CTCGATTTTGAAGCGGAGGTTTGAACCGCAGCCGCCATGCCGGGTTTACGAGATTG 342  
 QY 357 TGATCAAGGTGCCGAGCGACTGACGAGCACCCTGCCGGCATTTCTGACTCGTTGTGA 416  
 Db 343 TGATTAAGGTCCCGACGACCTTGACGAGCATCTGCCCGCATTTCTGACAGCTTGTGA 402  
 QY 417 GCTGGTGGCCGAGAGGAATGGAGCTGCCCGCGATTTGACATGATCTGAATCTGA 476  
 Db 403 ACTGGGTGGCCGAGAGGAATGGAGTTGCCCGCAGATTCTGACATGATCTGAATCTGA 462  
 QY 477 TTGAGCAGGCAACCCCTGACCGGTGCCGAGAAAGCTGCAGCGCGCACTTCTGTCCAATGCC 536  
 Db 463 TTGAGCAGGCAACCCCTGACCGGTGCCGAGAAAGCTGCAGCGCGCACTTCTGTCCAATGCC 522  
 QY 537 GCCCGGTGAGTAAGGCCCGGAGCCCTCTTCTTTGTTCACTTGCAGAGGGCGAGTCT 596  
 Db 523 GCCGTGTAGTAAGGCCCGGAGCCCTTTCTTTGTGCAATTGAGAGGAGAGAGCT 582  
 QY 597 ACTTCCACCTCCATATTTCTGTGAGACACGAGGGGTCAATCCATGTGCTGGCGCT 656  
 Db 583 ACTTCCACATGACAGTGTCTGTGAGAAACACGAGGGGTCAATCCATGTGTTGGAGCTT 642  
 QY 657 TCCGTAGTCAAGTTAGGAGACAGCTGTGCAGACCATCTACCGCGGATCGAGCCGACCC 716  
 Db 643 TCCGTAGTCAAGTTAGGAGACAGCTGTGCAGACCATCTACCGCGGATCGAGCCGACCT 702  
 QY 717 TGCCCAACTGTTTCCGCGGTGACCAAGCGCTAATGCGCGGAGAGGGGGAACAAGTGG 776  
 Db 703 TGCCCAACTGTTTCCGCGGTGACCAAGCGCTAATGCGCGGAGAGGGGGAACAAGTGG 762  
 QY 777 TGAGAGAGTGTACATCCCACTAATCTCTGCCCCAAGCTCAGCCGAGCTGAGTGG 836  
 Db 763 TGAGAGAGTGTACATCCCACTAATCTCTGCCCCAAGCTCAGCTGAGTGG 822  
 QY 837 CGTGAATAACATGAGAGGATTAAGCGCCTGTTGAACCTGCGGAGCGCAACGCGC 896  
 Db 823 CGTGAATAATATGAAACAGATTAAGCGCCTGTTGAATCTCAGGAGCGTAACGCT 882  
 QY 897 TCGTGGCGCAGCAGCCTGACCCAGCTCAGCCAGACAGCAGAGCAACAAGAGATCTGA 956  
 Db 883 TCGTGGCGCAGCAGCCTGACCCAGCTCAGCCAGACAGCAGAGCAACAAGAGATCTGA 942  
 QY 957 ACCCAATTTGACGCGCCTGTATCCGGTCAAAAACCTCCGCGCTACATGAGCTGG 1016  
 Db 943 ATCCCAATTTGATGCGCGCTGTATCAGATCAAAAACCTCAGCCAGGTACATGAGCTGG 1002  
 QY 1017 TCGGTGCTGTGTGACCGGCGCATCACTCCGAGAGAGTGGATCCAGAGAGCAGG 1076  
 Db 1003 TCGGTGCTGTGTGACCGGCGCATCACTCCGAGAGAGTGGATCCAGAGAGCAGG 1062  
 QY 1077 CTTGTACATCTCTTCAACGCGCTTCCAACTCGCGGTCCAGATCAAGCGCTTGG 1136  
 Db 1063 CTTGTACATCTCTTCAATGCGGCTCCCACTCGCGGTCCAAATCAAGCTGCTTGG 1122  
 QY 1137 ACAATGCCGCGCAAGATCATGGCGTGAACCAATCCGCGCGCACTACTGTAGGCCCG 1196  
 Db 1123 ACAATGCCGCGCAAGATCATGGCGTGAACCAATCCGCGCGCACTACTGTAGGCCCG 1182  
 QY 1197 CTCCGCGCGGACATTAACCAACCGCATCTACCGCATCTGAGCTGAACGCGCTAG 1256  
 Db 1183 AGCCGTGAGGACATTTCCAGCAATCGATTATTAATTTTGAACCTAAGCGGTAGC 1242  
 QY 1257 AACCTGCTACGCGCGCTCCGCTTTCTGCGGTGGGCCGAGAAAGTTGCGGAGCGCA 1316  
 Db 1243 ATCCCAATATATGCGGCTTCCGCTTTCTGAGATGGGCCAGAAAGTTGCGGAGCGCA 1302  
 QY 1317 ACACCATCTGCTGTTTGGCGCGCAACACGCGGCAAGCAACATCGCGGAGCGCATCG 1376  
 Db 1303 ACACCATCTGCTGTTTGGCGCGCAACACGCGGCAAGCAACATCGCGGAGCGCATAG 1362  
 QY 1377 CCGAGCGCGTGCCTTTCTACGCGTGCCTCACTGACCAATGAGAACTTCCCTCAATG 1436  
 Db 1363 CCGACACTGTGCCCTTCTACGCGTGCCTCAACTGACCAATGAGAACTTCCCTCAACG 1422





Db	3571	CAGTCAAGGCTTCAGTTTCTCAGGCCGAGCGAGTGCATTCGGGACCAGTCTAGGAAC	3630
QY	3654	TGGTACCTGGAACCCCTGTATCCGACGACGCCGCTTCTAAACAAAAACAGACAAC	3713
Db	3631	TGGCTTCCTGGACCCCTGTATCCGCCACGACGAGTATCAAGACATCTCGGATACAAAC	3690
QY	3714	AACAGCAATTTTACCTGGACTGGTCTTCAAAATATTAACCTCAATGGGCGGTGAATCCATC	3773
Db	3691	AACAGTGAATACTCGTGGACTGGAGCTACCAAGTACCCTCAATGGCAGAGACTCTCTG	3750
QY	3774	ATCAACCCCTGGCACTGTCTATGGCCTCACACAAGACGACGAAGCAAGTTCTTCCCATG	3833
Db	3751	GTGAATCCGGCCCGCCCATGGCAAGCCACCAAGGACGATGAAGAAAGTTTTCCTCAG	3810
QY	3834	AGCGGTGTCATGATTTTGGAAAAAGAGAGCGCCGAGCTTCAAACTGCTATTGGCAAT	3893
Db	3811	AGCGGGGTTCATCTTGGGGAAGCAAGGCTCAGAGAAAAAACAATGTGACATTGAAGAAG	3870
QY	3894	GTCAATGATTACAGACGAAGAAATTAAGCCACTAACCTGTGGCCACCGAAGATTT	3953
Db	3871	GTCAATGATTACAGACGAAGAAATCAGACACCAATCCCGTGGCTACGGAGCATAT	3930
QY	3954	GGGACCGTGGCACTCAATTTCCAGACGACGACGACACCACTGCGACCCGAGATGCAAT	4013
Db	3931	GGTTCGTATCTACCAACCTCCAGAGAGGCAACAGACAAGCTACCGCAGATGTCAAC	3990
QY	4014	GCTATGGAGCATTAACCTGGCATGGTGTGGCAAGATAGACGTGTACCTGCAGGGTCC	4073
Db	3991	ACACAAGGCGTCTTCCAGGCAATGGTCTGGCAGGACAGAGATGTATCCTTCAGGGGCC	4050
QY	4074	ATTGGGCAAAATCTCTACAACAGATGACACTTTCACCCGTCTCTTATGGCGGC	4133
Db	4051	ATCTGGCAAAAGATTCACACACGACGACATTTTCAACCTCTCCCTCATGGGTGA	4110
QY	4134	TTTGACTCAAGAACCCGCTCTCAGATCCTCATCAAAAACAGCGCTGTCTGCGCAAT	4193
Db	4111	TTGGAATTAACACCCCTCTCCACAGATTTCTATCAAGAACACCCGCTACTGCGAAT	4170
QY	4194	CCTCCGGCGAGTTTCAAGCTACAAGTTTGTCTTCAATCACCAATACCTCCACAGA	4253
Db	4171	CCTTCGACCACCTTCAGTGGCGGCAAGTTTGTCTTCTTCAATCACACAGTACTCCACGGGA	4230
QY	4254	CAAGTAGTGGAAATTTGAATGGGAGCTGCAGAAAGAAAACAGCAAGCGCTGGAATCCC	4313
Db	4231	CAGGTCAAGCTGGAGATCGAGTGGGAGCTGCAGAAAGAAAACAGCAAGCGCTGGAATCCC	4290
QY	4314	GAAATGAGTACACATCCAAATTAATGCAAAATCTGCCAAGTGAATTTTACTGTGCAAC	4373
Db	4291	GAAATTCAGTACACTTCCAACTACAACAAGTCTGTAAATGTGACTTACTGTGCAACT	4350
QY	4374	AATGACTTTATAGTGAAGCTTGCACCATTTGGCACCCGTTAACCCTGCCGTAA	4433
Db	4351	AATGCGGTATTCAGAGCCTCGCCCATTTGGCACCAAGATACCTGACTCGTAATCTGTA	4410
QY	4434	TTACGTGTTAATCAATAAACCCGTTGATTGCTTCAAGTTGAACCTTGTCTCTGTCTT	4493
Db	4411	TTGCTTGTAAATCAATAAACCGTTTAATTCTGTTCAAGTTGAACCTTGTCTCTGTCTT	4470
QY	4494	CTTATCTTATCGGTTACATGCTTATAGCTTACACATTAACCTGCTTGTGCGCTTCCG	4553
Db	4471	TCTTCTTATCTAGTTTCCATGGCTACGTAGATTAAGTACATGGCGGTTAATCATTAAC	4530
QY	4554	ATAAAAGACTTACGTATCGGCTTAACCCCTAGTGAATGAGTTGCCACTCCCTCTCTGCG	4613
Db	4531	TACAAGGA-----ACCCTAGTGAATGAGTTGGCCACTCCCTCTCTGCG	4574
QY	4614	CGCTCGCTCGCTCGGTGGGCGCTGCGGACCAAAAGTCCGACGCGCAGAGCTCTGCTCT	4673
Db	4575	CGCTCGCTCGCTCACTGAGCGCGGCGACCAAAAGTCCGCGGCGCGGCTTTCGCG	4634
QY	4674	GCCGCGCCCAACCGAGCGAGCGCGCAGAGAGGAGTGGGCAA	4718

```

Db      4635 GGGCGCCTCAGTACGACGCGGCGGCGGAGAGAGGAGTGGCCAA 4679

RESULT 6
AAF23750
ID      AAF23750 standard; DNA; 4679 BP.
XX
AC      AAF23750;
XX
DT      28-MAR-2001 (first entry)
XX
DE      AAV2 DNA sequence.
XX
KW      AAV2; gene therapy; AAV viral vector; cystic fibrosis; cancer; AIDS;
KW      atherosclerosis; sickle cell anaemia; thalassaemia;
KW      blood clotting disorder; diabetes; ss.
XX
OS      Adeno associated virus.
XX
PN      US6156303-A.
XX
PD      05-DEC-2000.
XX
PF      11-JUN-1997; 97US-0873168.
XX
PR      11-JUN-1997; 97US-0873168.
XX
PA      (UNIW ) UNIV WASHINGTON.
XX
PI      Russell DW, Rutledge EA;
XX
DR      WPI; 2001-060164/07.
XX
PT      Adeno-associated virus serotype 6 and viral vector derived from it for
PT      gene therapy of cystic fibrosis, cancer, acquired immunodeficiency
PT      syndrome, sickle cell anemia, thalassemia and diabetes
XX
PS      Claim 7; Fig 1; 50pp; English.
XX
CC      The present invention relates to adeno-associated virus serotypes. The
CC      present sequence is the DNA sequence of one such serotype (AAV2). AAV2
CC      can be used to construct AAV viral vectors for use in gene therapy for a
CC      range of disorders: cystic fibrosis, cancer, AIDS, atherosclerosis,
CC      sickle cell anaemia, thalassaemia, blood clotting disorders and diabetes.
CC      The AAV viral vectors have increased transduction efficiency of a
CC      particular host cell as the AAV virion containing the AAV vector genome
CC      can be modified to express a capsid protein of an AAV serotype that
CC      transduces the selected host cell.
XX
SQ      Sequence 4679 BP; 1197 A; 1263 C; 1255 G; 964 T; 0 other;

Query Match          65.8%; Score 3106.2; DB 22; Length 4679;
Best Local Similarity 80.2%; Pred. No. 0;
Matches 3789; Conservative 0; Mismatches 883; Indels 53; Gaps 10;

QY      1 TTGCCACTCCCTCTTCGCGCGCTCGCTCGTCCGTGGGCGCTGCGGACCAAGTCCGC 60
DB      1 TTGGCACTCCCTCTCTGCGCGCTCGCTCGTCACTGAGCGCGGCGGACCAAGTCCGC 60
QY      61 AGACGGCAGAGCTCTGCTCTGCGCGCGCTCGCTCGTCCGTGGGCGCTGCGGACCAAGTCCGC 120
DB      61 CGACGCGCGGCTTTGCGCGCGCGCTCGCTCACTGAGCGGAGCGGCGGAGAGGAGTGC 120
QY      121 GGCACCTCCATCACTAGGGGTAATCGCGAAGCGCTCCACGCTCGCGCTCAGCGCTGA 180
DB      121 GCCAACTCCATCACTAGGGGT-----TCTGAGGGGTGAGTGTGA 163
QY      181 CGTAAATTACGTCATAGGG---GAGTGTCTCTGTAATTAGCTGTACGAGTGTCTTTGC 237
DB      164 CGTGAATTACGTCATAGGGTTAGGAGGTCTGTATTAGAGTCACTGAGTG-TTTGC 222
QY      238 GACATTTTGCGACACGAGTGGCCATTTAGGGTATATAGCCGAGTGAGCGAGAGAT 297
DB      238 GACATTTTGCGACACGAGTGGCCATTTAGGGTATATAGCCGAGTGAGCGAGAGAT 297

```

Db 223 GACATTTTGGCAACCATGTGTCAAGCTGGGTATTTAAGCCCGAGTGACACGAGGT 282  
Qy 298 CTCATTTTGG-AACCGCAATTTGACGAGACGACCATGCGGGCTTCTACGAGATCG 356  
Db 283 CTCATTTTGAAGCGGGAGGTTGAACGCGCAGCCGCATGCGGGGTTTACGAGATTG 342  
Qy 357 TGATCAAGGTCCCGACCGACCTGACGAGCACTGCGGGCATTTCTGACTGTTTGTGA 416  
Db 343 TGATTAAGGTCCCGACCGACCTTGAACGGGCATCTGCCCCGCATTTCTGACAGCTTTGTGA 402  
Qy 417 GCTGGGTGGCCGAGAAAGGAATGGAGCTGCCCCGGGATTTCTGACATGATCTGAATCTGA 476  
Db 403 ACTGGGTGGCCGAGAAAGGAATGGAGTTGCCGCAGATTCTGACATGATCTGAATCTGA 462  
Qy 477 TTGAGCAGGCAACCCCTGACCGTGGCCGAGAAAGCTGACGCGCATTTCTGCTCAATGCG 536  
Db 463 TTGAGCAGGCAACCCCTGACCGTGGCCGAGAAAGCTGACGCGCATTTCTGACGGAATGGC 522  
Qy 537 GCCGCTGAGTAAGGCCCCGAGGCCCCCTTTCTTGTTCAGTTTCAAGTGAAGGCGAGTCCCT 596  
Db 523 GCCGTGTGAGTAAGGCCCCGAGGCCCCCTTTCTTGTGCAATTTGAGAAAGGAGAGAGCT 582  
Qy 597 ACTTCCACCTCCATATTCTGTGTGAGACCAACGCGGGTCAATCCATGTTGTGGGCCGCT 656  
Db 583 ACTTCCACATGACGTCGTCTCGTGAAACCAACGCGGGTGAATCCATGTTTGGGACGTT 642  
Qy 657 TCCTGAGTCAAGATTAGGGAACAAGCTGTGCAACCATCTACCGCGGGATCGAGCCGACCC 716  
Db 643 TCCTGAGTCAAGATTAGGGAACAAGCTGTGCAAGAAATTTACCGCGGGATCGAGCCGACTT 702  
Qy 717 TGCCCAACTGTTGCGGGTGACCAAGACGCGTAATGGCGCCGAGAGGGGGGAACAAGTGG 776  
Db 703 TGCCAAACTGTTGCGGGTCACAAAAGACCAAAATGGCGCCGAGAGGGGGGAACAAGTGG 762  
Qy 777 TGACGAGTGTACATCCCAACTACCTCTGCCCCAAGACTCAGCCCGAGCTGCAGTGG 836  
Db 763 TGATGAGTGTACATCCCAATTACTGTCTCCCAAAACCCAGCCGTGAGTCCAGTGG 822  
Qy 837 CGTGACTAACTAGAGAGATATATTAAGCGCTGTTGAACCTGCGCGGACCAAAACGCGC 896  
Db 823 CGTGACTAACTAGAGACAGTATTTAAGCGCTGTTGAATCTCAGCGAGCTTAAACGCT 882  
Qy 897 TCGTGCGCAGCACTGACCCACGTCAGCAGCAACCGAGAGAGAAACAAGAAATCTGA 956  
Db 883 TCGTGCGCAGCACTGACGACGTCGTCAGACGAGAGAGAGAAACAAGAAATCTGA 942  
Qy 957 ACCCCAATTCTGACGCGCTGTCTATCCCGTCAAAAACCTCCGCGGCTACATGAGCTGG 1016  
Db 943 ATCCCAATTCTGATGCGCGCTGTATCAGATGAATAAATTCAAGCAGGTACATGAGCTGG 1002  
Qy 1017 TCGGGTGGCTGTGACCGGGGATCACTCCGAGAAAGCAGTGGATCCAGAGGACCAAG 1076  
Db 1003 TCGGGTGGCTGTGACAAAGGGGATTAACCTCGAGAAAGCAGTGGATCCAGAGGACCAAG 1062  
Qy 1077 CCTGTACATCTCTTCAACGCGCTTCCAACTGCGGTCGCCAGATCAAGCGCTGTGG 1136  
Db 1063 CCTCATACATCTCTTCAATGCGGCTTCCAACTGCGGTCGCCAAATCAAGGCTGCTTGG 1122  
Qy 1137 ACAATGCGGCAAGATCATGCGCTGACCAAAATCGCGCCGACTACCTGTGTAAGCCCG 1196  
Db 1123 ACAATGCGGAAAGATATGAGCCTGACTAAACCGCCCCGACTACCTGTGTTGGCCAGC 1182  
Qy 1197 CTCCGCGCGGCAATTAACCAACCGCATCTACCGCATCTGAGCTGAACGGCTAGC 1256  
Db 1183 AGCCCGTGAGGACATTTCCAGCAATCGGATTTATAAATTTTGAACATAACGGGTAGC 1242  
Qy 1257 AACCTGCTACGCGCGCTCCGTCTTCTCGGCTGGCCCCAGMAAAGTTTCGGGAAGCGCA 1316  
Db 1243 ATCCCAATATGCGGCTCCGTCTTCTGAGATGGGCCACGMAAAGTTTCGGCAAGAGGA 1302  
Qy 1317 ACACCATGTGGCTGTTTGGGGCCGCGCAACCAAGGAGAGACCAACATCGCGGAGGCCATCG 1376  
Db 1303 ACACCATGTGGCTGTTTGGGGCTGCAACTACCGGAGAGACCAACATCGCGGAGGCCATAG 1362

Qy 1377 CCCACGCGGTGCCCTTCTACGGCTGCGTCAACTGGAACCAATGAGAACTTCCCTTCAATG 1436  
Db 1363 CCCACACTGTGCCCTTCTACGGGTGCGTAACTGGAACCAATGAGAACTTCCCTTCAACG 1422  
Qy 1437 ATTGCTGCAACAAGATGTGATCTGTGGGAGAGGAGGCAAGATGACGGCCAAAGTCTGG 1496  
Db 1423 ACTGTGTCACAAGATGTGATCTGTGGGAGAGGAGGAGGAAAGATGACCGCCAAAGTCTGG 1482  
Qy 1497 AGTCCGCAAGGCCATTCTCGCGGCAACAAGTGCAGTGAACCAAAAGTCAAGTCTG 1556  
Db 1483 AGTCCGCAAGGCCATTCTCGGAGAGCAAGGTGCAGTGAACCAAAATGCAAGTCTG 1542  
Qy 1557 CCGCCGATCGACCCCAACCCCGTGTACCTCAACCAACCAATGTGCGCGCTGA 1616  
Db 1543 CCGCCGATAGACCCGACTCCCGTGTGTCACCTCAACCAACCAATGTGCGCGCTGA 1602  
Qy 1617 TTGACGGGAACAGCACCACTTCAGACACCAAGCAGCCGTTGACAGACCGGATGTTCAAT 1676  
Db 1603 TTGACGGGAACCAACGACCTTCAGACACCAAGCAGCCGTTGACAGACCGGATGTTCAAT 1662  
Qy 1677 TTGAATCAACCCCGCTGTGAGACATGACTTTGGCAAGGTGACAAAGCAGGAAGTCAAG 1736  
Db 1663 TTGAATCAACCCCGCTGTGATCATGACTTTGGGAAGGTGACCAAGCAGGAAGTCAAG 1722  
Qy 1737 AGTTCCTCGCTGGGCGCAGATCACTGACCGAGTGGCGCATGAGTTCTACGTCAAG 1796  
Db 1723 ACTTTTCCGTTGGGCAAGGATCACTGTGTGAGGTGAGCATGAAATTTACGTCAAAA 1782  
Qy 1797 AGGTGAGCCAAACAAGAGACCCGCCCCGATGACGCGGATAAAGCAGGCCCAAGCGGG 1856  
Db 1783 AGGTGAGCCAAAGAGAGACCCGCCCCGATGACGCGGATAAAGCAGGCCCAAGCGGG 1842  
Qy 1857 CCTGCCCTCAGTCGCGGATCACTGACGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1916  
Db 1843 TGCGCGAGTCACTGCGAGCGCATGACGTCAAGAGCGCGA--AGCTTCGATCACTAGC 1899  
Qy 1917 CCGACAGGTACCAAAACAATGTTCTCGTCAACGCGGAGATGTTCAAGTGTGTTCCCT 1976  
Db 1900 CAGACAGGTACCAAAACAATGTTCTCGTCAAGTGTGAGTGTGTTCCCT 1959  
Qy 1977 GCAAGACATCGAGAGAGATGAATGAATTTCAACATTTGCTTCAACGAGAGAGAGAG 2036  
Db 1960 GCAAGACATCGAGAGAGATGAATGAATTTCAACATTTGCTTCACTCAAGAGAGAGAG 2019  
Qy 2037 ACTGTTCAAGTGTCTTCCCGCGGTGTCAGATCTCAACCGGT--CGTCAAGAAAGAGA 2093  
Db 2020 ACTGTTCAAGTGTCTTCC--CGTCAAGATCTCAACCGGTGCTGCTCAAAAAGG 2076  
Qy 2094 CGTATCGGAACTCTGTGCAATTCATCTGCTGGGGCGGGCTCCGAGATGCTTGTCT 2153  
Db 2077 CGTATCGGAACTCTGTCTACATTCATATCATGGAAGAGTGCAGAG--CGCTTGCA 2133  
Qy 2154 CGGCTGCGATCTGTCAACGTGACCTGATGACTGTGTTCTGACCAATAATGACTT 2213  
Db 2134 CTGCTGCGATCTGTCAATGTGATTTGGATGACTGATCTTTGAACAATAATGATTT 2193  
Qy 2214 AAACCAAGTATGGCTGCCGATGTTATCTTCCAGATTTGGCTCGAGACAACCTCTGAG 2273  
Db 2194 AAATCAGTATGGCTGCCGATGTTATCTTCCAGATTTGGCTCGAGACAACCTCTGAG 2253  
Qy 2274 GGCATTCGGAAGTGTGGGACTTGAACCTGGAAGCCCGAAGCCCAACCAAGCA 2333  
Db 2254 GGAATTAAGCAGTGTGGAAGCTCAACCTGGCCCAACCAACCAAGCCCGAAGCGG 2313  
Qy 2334 AAGCAGAGAGAGCGCGGGGTCTGTGCTTCTGCTACAGTAACTTGGACCTTCAAC 2393  
Db 2314 CATTAAGAGAGAGAGGGGTCTGTGCTTCTGCTTCAAGTAACTTGGACCTTCAAC 2373  
Qy 2394 GGAATTCGGAAGTGTGGGACTTGAACCTGGAAGCCCGAAGCCCAACCAAGCA 2453  
Db 2374 GGAATTCGGAAGTGTGGGACTTGAACCTGGAAGCCCGAAGCCCAACCAAGCA 2433



QY 2454 GCCTACGACGAGCTCAAAGCGGGTGACATCCGTACCTGCGGTATTAACCAAGCCGAC 2513  
|||  
Db 2434 GCCTACGACCGGAGCTCGACAGCGGAGCAACCGTACCTCAAGTAAACCAAGCCGAC 2493  
QY 2514 GCCGAGTTTCAGAGCGCTCTGCAAGAAATACGTTTGGGGCAACTCGGCGAGCA 2573  
|||  
Db 2494 GCGAGTTTCAGAGCGCTTAAAGAAATACGTTTGGGGCAACTCGGCGAGCA 2553  
QY 2574 GTCTTCAGGCGCAAGAGCGGTTCTGAACTCTCGTCTGTTGAGAAAGCGCTAAG 2633  
|||  
Db 2554 GTCTTCAGGCGCAAGAGAGGTTCTGAACTCTGGGCTGTTGAGAACTGTAAAG 2613  
QY 2634 ACCGCTCCTGAAAGAAACGTCGCGTACAGAGTCCGCACAAGAGCAGACTCCTCTCG 2693  
|||  
Db 2614 ACCGCTCCGGGAAAAAAGAGCGCGGTACAGCACTCTCTGTGAGCCAGACTCCTCTCG 2673  
QY 2694 GGCATCGGCAAGACAGGCCAGAGCCCGCTAAAAAGAGACTCAATTTTGTGAGACTGGC 2753  
|||  
Db 2674 GGAACCGGAAAGCGGGCCAGAGCTTGCAAGAAAAAGATTGAATTTGTGAGACTGGA 2733  
QY 2754 GACTCAGAGTCACTCCCGATCCAACTCTCGGAGAACTTCAGAAACCCCGCTGCT 2813  
|||  
Db 2734 GACGCAAGTCACTACCTGACCCCGAGCTCTCGGACAGCAAGCAGACCCCTCTGCT 2793  
QY 2814 GTGGAGCTACTCAATGCTTCAAGCGGTGGCGGACCAATGGCAGACAATAAGAGGC 2873  
|||  
Db 2794 CTGGGAATTAATACGATGGCTACAGGAGTGGCGGACCAATGGCAGACAATAAGAGGC 2853  
QY 2874 GCCGACGAGTGGGTAATGCTCTCAGAAATTGGCATTTGCCATTGCGATGGCTGGCGAC 2933  
|||  
Db 2854 GCCGACGAGTGGGTAATTCCTCCGAAATTGGCATTTGCCATTGCGATGGCGAC 2913  
QY 2934 AGAGTCAATCAACACGACCCCGACCTGGGCTTGCCCACTCAATAACCACTCTAC 2993  
|||  
Db 2914 AGAGTCAATCAACACGACCCCGAAGCTGGGCTTGCCCACTCAATAACCACTCTAC 2973  
QY 2994 AAGCAAAATCTCAGTCTTCAACGGGGCCAGCAAGCAACCACTACTTGGCTACAGC 3053  
|||  
Db 2974 AAACAAATTTCCAGCCAATCA--GAGGCTCGAAGCAATCACTTTGGCTACAGC 3030  
QY 3054 ACCCCCTGGGGTATTTGATTTCAAGATTCACCTGCACTTTTCAACCACTGACTGG 3113  
|||  
Db 3031 ACCCCCTGGGGTATTTGACTTCAAGATTCACCTGCACTTTTCAACCACTGACTGG 3090  
QY 3114 CAGGACTCATCAACACATTTGGGATTCGGGCCAAGAGACTCACTTCAACTCTTC 3173  
|||  
Db 3091 CAAAGACTCATCAACACACTGGGGATTCGCAACCAAGAGACTCAACTTCAAGCTCTTT 3150  
QY 3174 AACATCCAAGTCAAGGAGTCAAGCAATGATGGCGTCAACCACTGCTAATAACTTT 3233  
|||  
Db 3151 AACATCCAAGTCAAGAGGTCAAGCAATGACGTAAGCAAGATTCCTAATAACTTT 3210  
QY 3234 ACCAGCAGGTTCAAGTCTTCTCGGACTCGGAGTACCAAGCTTCGTAAGTCTCGCTCT 3293  
|||  
Db 3211 ACCAGCAGGTTCAAGTCTTACTGACTCGGAGTACCAAGCTTCGTAAGTCTCGCTCTG 3270  
QY 3294 GCGCAGCAGGGCTGCTCCCTCGTTCCGGCGGAGAGTGTTCATGATTCGCAATAAGGC 3353  
|||  
Db 3271 GCGCATCAAGAGATGCTCCCGCGTTCCAGCAGAGAGTCTTCAATGTTGCAAGTATGA 3330  
QY 3354 TACCTGACGCTCAACATGGCAGCCAGCCGTGGAGCTTCATCTTTTACTGCTGGA 3413  
|||  
Db 3331 TACCTGACCTGAACAAAGGAGTCAAGCAATGAGACGCTCTTCAATTTTACTGCTGAG 3390  
QY 3414 TATTTCCCTTCTCAGATGCTGAGAAAGGCAACAACTTTACTTTCAGCTACACTTTGAG 3473  
|||  
Db 3391 TACTTCTTCTCAGATGCTGCTGACCGGAAACAACTTTACTTTCAGCTACACTTTGAG 3450  
QY 3474 GAAGTCCCTTTCACAGCAGCTACGCGACAGCCAGAGCTGAGCCGGCTGATGAATCT 3533  
|||  
Db 3451 GACGTTCTTTCACAGCAGCTACGCTCAAGCCAGAGTCTGAGCCGTCTCATGAATCT 3510  
QY 3534 CTATGACCAATATCTGTATTAATTAAGAACTCAAAATCAGTCCGGAAGTGCCCA 3593

Db 3511 CTATGACCAAGTACCTGTATTAATTTAGCAGAGAAACAACTCCAAAGTGAACCAACGACG 3570  
QY 3594 AACAGGACTTGCTGTTAGCCGTGGGTCTCCAGCTGGCATGTCTGTTAGCCCAAAAC 3653  
|||  
Db 3571 CAGTCAAGGCTTCAAGTTTCTCAGGCGCGAGCGAGTGAATTCGGGACCAAGTCTAGAAC 3630  
QY 3654 TGGCTACTGAGCCCTGTTATCGGACAGCGCGGTTTCTAAAACAAAAACAGACAAC 3713  
|||  
Db 3631 TGGCTTCTGAGCCCTGTATCCGACAGCGAGTATCAAGACATCTGCGATTAACAC 3690  
QY 3714 AACAGCAATTTTACCTGAGCTGCTTCAAAATATTAACCTCAATGGGCGTGAATCAATC 3773  
|||  
Db 3691 AACAGTGAATACTCGTGAGCTGAGCTACCAAGTACCACTCAATGGCAGAGACTCTCTG 3750  
QY 3774 ATCAACCTGGCACTGCTATGCGCTCACACAAGACGACGACAGCAAGTCTTCCCATG 3833  
|||  
Db 3751 GTGAATCCGGGCGCGCCATGGCAAGCCACAAGACGATGAAGAAAGTTTCTCTCAG 3810  
QY 3834 AGCGGTCTCATGATTTTGGAAAAAGAGCGCGGAGCTTCAACACTGCAATGGACAT 3893  
|||  
Db 3811 AGCGGGTCTCATCTTTGGGAAGCAAGGCTCAGAGAAAACAAAATGTGACATTTGAAG 3870  
QY 3894 GTCATGATTAACAGCAGAGAAATTAAGCCACTAACCTGTGCGCACCGAAAGATT 3953  
|||  
Db 3871 GTCATGATTAACAGCAGAGAAATTAAGCAACCAATCCGTGCTACGAGCAGTAT 3930  
QY 3954 GGCACCGTGGCAGTCAATTTCCAGAGCAGACAGACACCTGCGACCGAGATGCAAT 4013  
|||  
Db 3931 GGTCTGTATCTACCAACCTCAGAGAGGCAACAGACAGCACTACCGAGATGTCAAC 3990  
QY 4014 GCTATGGAGCATTAACCTGGCATGTGTGGCAAGATAGACAGTGTACTGAGGTTCC 4073  
|||  
Db 3991 ACACAAGCGGTTCTTCCAGGATGTGTGGCAGGACAGAGATGTACTTACGGGCTCC 4050  
QY 4074 ATTGGGCCAAATTCCTCAACAGATGACACTTACCCGCTCTCTTATGCGGCGC 4133  
|||  
Db 4051 ATCTGGGCAAAATTCACACAGCAGACGACATTTTCAACCCCTCTCCCTCATGGGTGA 4110  
QY 4134 TTGGAATCAAGAACCGGCTCTCAGATCCTCATCAAAAAACAGCCTGTTCTCGAAT 4193  
|||  
Db 4111 TTGGAATTAACACACCTCTCTCAAGATTTCTCATCAAGAACACCCGCTACTCGAAT 4170  
QY 4194 CTTCCGGCGAGTTTCAAGCTAACAAAGTTGCTTCAATCATCAACCAATACTCCACAGA 4253  
|||  
Db 4171 CTTCCGACCACTTCAAGTGGGCAAGTTGCTTCTCTCATCAACAGTACTCCAGGGA 4230  
QY 4254 CAAGTGAAGTGAATTTGAATGGAGCTGCAGAAAGAAAAACAGCAAGCGCTGAATCC 4313  
|||  
Db 4231 CAGGTCAAGCTGAGATGAGTGGAGCTGCAGAAAGAAAAACAGCAAGCGCTGAATCC 4290  
QY 4314 GAAATGCAATACATCCAAATTAATGCAAAATCTGCCAAGTGAATTTACTGTGACAA 4373  
|||  
Db 4291 GAAATTCAGTACACTTCAACTAACAAAGTGTGAATGTGAATTTACTGTGACAACT 4350  
QY 4374 AATGGAATTTATAGAGCTCGCCCATTTGGCAGCCGTTAACCCTCCCTGTAA 4433  
|||  
Db 4351 AATGGCTGTATTAGAGCTCGCCCATTTGGCAGCAATACCTGACTGTAACTGTAA 4410  
QY 4434 TTAAGTGTAAATCAATAACCGGTTGATTCGTTTCAAGTTGAATTTGCTCTCTCT 4493  
|||  
Db 4411 TTGCTGTAAATCAATAACCGTTAATTCGTTCAAGTTGAATTTGCTCTCTCTCT 4470  
QY 4494 CTATCTTATCGGTTACATGTTATAGCTTACATTAATCTGCTGCTGCTGCGG 4553  
|||  
Db 4471 TCTTCTTATCTAGTTTCCATGGCTACGTAGATAGTACATGGCGGTTAATCATTAAC 4530  
QY 4554 AATAAAGACTTACGTCATCGGGTTACCCCTAGTGAAGTTGCCCACTCCCTCTGCG 4613  
|||  
Db 4531 TACAAGG-----ACCCCTAGTGAAGTTGGCCACTCCCTCTCTGCG 4574  
QY 4614 CGCTCGCTCGCTCGGTGGGCTTGCGAACCAAGTCCGACAGCGGAGAGCTCTGCTCT 4673  
|||



Db 1123 ACAATGGGGAAAGATTATGAGCTGACTAAACCGCCCCGAGCTACTGTGGGCCAGC 1182  
QY 1197 CTCGCCCGCGGACATTAAACCAACCGCATCTACCGCATCTGGAGCTGAACGGCTACG 1256  
Db 1183 AGCCCGTGGAGGACATTTCAGCAATCGGATTATATAATTTTGGAACCTAAACGGGTACG 1242  
QY 1257 AACCTGCTTACGCCGGCTCCGCTTTTCTCGGCTGGGCCAGAAAAGGTTGGGAAGCGCA 1316  
Db 1243 ATCCCAATATGCGGCTCCGCTTTTCTGGGATGGGCCAGAAAAGTTTCGGCAAGAGGA 1302  
QY 1317 ACACCATCTGGCTGTTGGGCGGCGCACCGGGCAAGACCAATCGCGGAAGCCATCG 1376  
Db 1303 ACACCAATCTGGCTGTTGGGCTGCACTACCGGAAGACCAATCGCGAGGCCATAG 1362  
QY 1377 CCGACGCGTGCCCTTCTACCGGCTGCGTCAACTGACCAATGAACTTCCCTTCAATG 1436  
Db 1363 CCGACACTGTGCCCTTCTACGGGTGCGTAAACTGACCAATGAACTTCCCTTCAACG 1422  
QY 1437 ATTCGCTGACACAAGATGTGATCTGTGGGAGGAGGGCAAGATGACGGCCAAAGTCTGG 1496  
Db 1423 ACTGTGTGACAAGATGTGATCTGTGGGAGGAGGGGGAAGATGACCGCCAAAGTCTGG 1482  
QY 1497 AGTCGCGCAAGGCCATTCTCGGCGGACGAAGGTGCGGCTGACCAAAAGTGCAGTCTG 1556  
Db 1483 AGTCGCGCAAAGCCATTCTCGAGGAAAGCAAGGTGCGGCTGACCAAAATGCAGATCT 1542  
QY 1557 CCGCCAGATCGACCCCGACCCCGGTGATCGTCACTCCAAACCAATGTGCGCCGTGA 1616  
Db 1543 CGGCCAGATAGACCCGACTCCCGTGTGATCGTCACTCCAAACCAATGTGCGCCGTGA 1602  
QY 1617 TTGACGGGAACAGACCACTTCGAGCACGACGACCGGTTGCAAGACCGGATGTTCAAT 1676  
Db 1603 TTGACGGGAACCTCAACGACCTTCGACACGACGACCGGTTGCAAGACCGGATGTTCAAT 1662  
QY 1677 TTGAACCTCACCCCGCGCTGTGAGCATGACTTTGGCAAGGTGACAAAGCAGAGAGTCAAG 1736  
Db 1663 TTGAACCTCACCCCGCGCTGTGATCATGACTTTGGGAAGGTCAACAGCAGAGAGTCAAG 1722  
QY 1737 AGTTCCTCCGCTGGGCGCAGGATCACGTGACCGAGGTGGCCGATGAGTTCTACGTCAGAA 1796  
Db 1723 ACTTTTCCGGTGGGCAAAAGGATCACGTGTTGAGGTGAGCATGAATTTCTACGTCAGAA 1782  
QY 1797 AGGCTGAGCCAAACAAAAGACCCGCCCCGATGACGCGGATAAAGCGAGCCCCAAGCGGG 1856  
Db 1783 AGGCTGAGCCAAAGAAAAGACCCGCCCCGATGACGAGATTAAGTGAAGCCCAACGGG 1842  
QY 1857 CCTGCCCCCTCAGTCGCGGATCCATGACGTGACGACGCGGAAGAGCTCCGCTGACTTTG 1916  
Db 1843 TGCGGAGTCAAGTTGCGGACGCACTGACGTGACGCGGA--AGCTTCGATCAACTACG 1899  
QY 1917 CCGACAGGTACCAAAACAATGTTCTCGTCAACGCGGACATGCTTCAGATGCTGTTCCCT 1976  
Db 1900 CAGACAGGTACCAAAACAATGTTCTCGTCAACGCGGACATGATGCTGTTCCCT 1959  
QY 1977 GCAAGACATGCGAGAGATGAATCAGAAATTTCAACATTTGCTTCAACGACGGAACGAG 2036  
Db 1960 GCAGACAATGCGAGAGATGAATCAGAAATTTCAAAATCTGCTTCACTCACGGAACAGAAAG 2019  
QY 2037 ACTGTTCAAGAGTCTTCCCGCGGTGTCAGAAATCTCAACCCGCT--CGTCAGAAAGAGGA 2093  
Db 2020 ACTGTTTGAAGTCTTTC--CGTTCAGAAATCTCAACCCGTTTCTGTCTCAAAAAAG 2076  
QY 2094 CGTATCGAAAACTCTGTGCCATTCATCTGTGCGGGCGGCTCCGAGATTGCTTCT 2153  
Db 2077 CGTATCAGAAACTGTCTACATTCATATATCATGGAAGGTGCCAGA--CGCTTGA 2133  
QY 2154 CGGCTGCGATCTGTCAACGTGACCTGATGACTGTGTTCTGAGCAATAATGACTT 2213  
Db 2134 CTGCTCGCATCTGTCAATGTGATTTGGATGACTGCATCTTTGAACAATAATGATTT 2193  
QY 2214 AAACCAAGTATGCTGCCGATGTTATCTTCCAGATTGGCTCGAGGACCAACTCTCTGAG 2273

Db 2194 AAATCAGTATGCTGCCGATGCTGTTATCTTCCAGATTGGCTCGAGGACACTCTCTGAA 2253  
QY 2274 GGCATTGCGAGTGTGGGACTTGAAAACCTGAGCCCCGAAAGCCCAAGCCAAACGACAA 2333  
Db 2254 GGAATTAAGACAGTGTGGAAGCTCAAAACCTGGGCCACCAACCAAGCCCGAGAGCGG 2313  
QY 2334 AAGCAGACGACGCGCGGGGTCTGTGCTTCTGCTGCTCAAGTAACTCCGACCTTCAAC 2393  
Db 2314 CATTAAGACGACAGAGGGGTCTGTGCTTCTGGGTACAAGTAACTCCGACCTTCAAC 2373  
QY 2394 GGACTCGACAGGGGAGCCCGTCAACGCGGCGGACCGACGGCCCTCGAGCACGACAAAG 2453  
Db 2374 GGACTCGACAGGGGAGCCCGTCAACGAGGCGACGCGCGGCCCTCGAGCACGACAAA 2433  
QY 2454 GCCTACGACGACGCTTCAAGCGGGTGAACAATCGGTACTTCGGTATAACGACGCTGAC 2513  
Db 2434 GCCTACGACCGGACGCTGACAGCGGAGACCAACCGTAACTCAAGTACAAACGACGCGAC 2493  
QY 2514 GCGGATTTCAAGACGCTGTCAGAAAGATACGTTTGGGGCAACCTCGGGGAGCA 2573  
Db 2494 GCGGATTTCAAGACGCGCTTAAAGAAAGATACGTTTGGGGCAACCTCGGACGACGA 2553  
QY 2574 GTCTTCCAGGCGCAAGACGGGTTCTCGAACCTCTCGGTCTGTTGAGGAAGCGCTAAG 2633  
Db 2554 GTCTTCCAGGCGAAAGAGGGGTTCTTGAACCTCTGGGCTGTGAGGAACCTGTTAAG 2613  
QY 2634 ACGGCTCTGGAAGAAACGTCGCGTAGAGCAGTCGCCACAAGACCAAGCTCTCTCTCG 2693  
Db 2614 ACGGCTCGGGAAMAAAAGAGCCCGTAGAGCAGCTCTCTGTGAGACCAAGCTCTCTCG 2673  
QY 2694 GGCAATCGGCAAGACAGGCCAGCAGCCCGCTAATAAAGAGACTCAATTTTGTGACACTGGC 2753  
Db 2674 GGAACCGGAAGCGGGGCGAGCAGCCTGCAAGAAAAGATTGAATTTTGTGACACTGGA 2733  
QY 2754 GACTCAGAGTCAATCCCGATCCACAACTCTCGGAGAACTTCAGCAACCCCGCTGCT 2813  
Db 2734 GACGCAAGACTCAGTACTGACCCCGAGCCTCTCGGACAGCCACGACAGCCCGCTCTGCT 2793  
QY 2814 GTGGACCTACTACAAATGCTTCAAGCGGTGGCGGACCAATGGCAGACAAATAAGAAAGC 2873  
Db 2794 CTGGGAACATAAGATGCTTACAGGACGTGGCGCAATGGCAGACAAATAAGAAAGC 2853  
QY 2874 GCCGACGAGTGGGTAAATGCTCAGGAATTTGGCATTTGCCATTCCACATGCGTGGCGAC 2933  
Db 2854 GCCGACGAGTGGGTAAATTCCTCCGGAATTTGGCATTTGCCATTCCACATGATGGCGAC 2913  
QY 2934 AGAGTCATCACCAACGACACCCGACCTGGGCTTGGCCCACTCAATAACCACTCTAC 2993  
Db 2914 AGAGTCATCACCAACGACACCCGACCTGGGCTTGGCCCACTCAATAACCACTCTAC 2973  
QY 2994 AAGCAATCTCCAGTGTCTTCAACGGGGCGCAGCAACGACCACTACTTGGCTACAGC 3053  
Db 2974 AAACAAATTTCCAGCCAATCA--GGAGCTTCGAACGACAAATCACTACTTGGCTACAGC 3030  
QY 3054 ACCCTCTGGGGTATTTGATTTCAACAGATTCGACCTTTCACCACTGACTGG 3113  
Db 3031 ACCCTTGGGGTATTTGACTTCAACAGATTCGACCTTTCACCACTGACTGG 3090  
QY 3114 CAGGACTCATCAACAACAATTGGGATTCGCGCCCAAGAGACTCAACTTCAACTCTTC 3173  
Db 3091 CAAGAAGTCAACACAACTGGGATTCGACCAAGAGACTCAACTTCAAGCTCTT 3150  
QY 3174 AACATCCAAGTCAAGAGGTCAACGACGATGATGCGTCAACAACATCGCTAATACTT 3233  
Db 3151 AACATCCAAGTCAAGAGGTCAACGACGATGATGCGTCAACAACATCGCTAATACTT 3210  
QY 3234 ACCAGCAGGTTCAAGTCTTCTCGGACTCGGAGTACCAAGCTTCCGTAAGTCTCGGCTCT 3293  
Db 3211 ACCAGCAGGTTCAAGTCTTCTCGGACTCGGAGTACCAAGCTTCCGTAAGTCTCGGCTCTG 3270  
QY 3294 GCGCACCAGGGCTGCTCCCTCGTTCCCGCGGACGTTGTCATGATTCGGAATAAGGC 3353  
Db 3271 GCGCATCAAGGATGCTCCCGCTTCCAGCAGAGCTCTTCAATGATGCGCACAGTATGA 3330



QY 3354 TACCTGACGCTCAACAATGGCAGCCGAGCGGTGAGCGTTCATCTTTTACTGCTGGA 3413  
 DB 3331 TACCTGACCGCTGAACAACGGGAGTACAGGACGAGTACGCTCTTCACTTTTACTGCTGAG 3390  
 QY 3414 TATTTCCCTTCAGATGCTGAGAACGGGCAACAATTACCTTACGCTACACCTTTGAG 3473  
 DB 3391 TACTTCCCTTCAGATGCTGAGAACGGGCAACAATTACCTTACGCTACACCTTTGAG 3450  
 QY 3474 GAAGTCCCTTCACAGACGCTACGCGCACAGCCGAGCGGTGAGCGGTGATGATCCT 3533  
 DB 3451 GACGTCCTTTCACAGACGCTACGCTACAGCGAGTGTGACCGGTCTCATGAATCCT 3510  
 QY 3534 CTGATGACCAATACCTGTATTACCTGAAGAACTCAAAATCAGTCCGGAAGTCCCAA 3593  
 DB 3511 CTGATGACCAATACCTGTATTACCTGAAGAACTCAAAATCAGTCCGGAAGTCCCAA 3570  
 QY 3594 AACAGGACTTGTGTTTACCGGTGCTCCAGCTGGCATGTCTGTGACCCCAAAAC 3653  
 DB 3571 CAGTCAAGGCTTCACTTTTCTCAGGCGCGAGCGAGTACATTCGGGACCACTAGGAAC 3630  
 QY 3654 TGGCTACCTGACCGCTGTATTCGGCAGCAGCGCGTTCCTTAAACAAACACAGACAAAC 3713  
 DB 3631 TGGCTTCCCTGACCGCTGTATTCGGCAGCAGCGCGTTCCTTAAACAAACACAGACAAAC 3690  
 QY 3714 AACAGCAATTTACTGGAAGTGTGCTTCAAAATATTAACCTCAATGGCGGTGAATCCATC 3773  
 DB 3691 AACAGCAATTTACTGGAAGTGTGGAAGTGTGGAAGTGTGGAAGTGTGGAAGTGTGGAAG 3750  
 QY 3774 ATCAACCTGGAAGTGTGGAAGTGTGGAAGTGTGGAAGTGTGGAAGTGTGGAAGTGTGGAAG 3833  
 DB 3751 GTGAATCCGGGCGCGCGCATGGAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 3810  
 QY 3834 AGCGGTGTCATGATTTTGGAAAAGAGAGCGCGGAGCTTCAACACATGTCATGGAACAT 3893  
 DB 3811 AGCGGTGTCATGATTTTGGAAAAGAGAGCGCGGAGCTTCAACACATGTCATGGAACAT 3870  
 QY 3894 GTCATGATTACAGACGAGAGGAATTAAGCCACTAACCTGTGCGCACCGAAGATTT 3953  
 DB 3871 GTCATGATTACAGACGAGAGGAATTAAGCCACTAACCTGTGCGCACCGAAGATTT 3930  
 QY 3954 GGGACCGTGGCAGTCAATTTCCAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGT 4013  
 DB 3931 GGGTCTGTATCTACCAACTCTCCAGAGAGGCAACAGACAGCAGCAGCAGCAGCAGCAGCAGCAG 3990  
 QY 4014 GCTATGGAGCATTAACCTGGCATGTGTGGCAAGATAGAGACGTGTACTGCAAGGCTCC 4073  
 DB 3991 ACACAAAGGCGTCTTCCAGGAGTGTGTGGCAGAGAGATGTGTACTTTCAGGGGCCCC 4050  
 QY 4074 ATTTGGCCAAATTCCTCACAAGATGACACTTTCACCCGCTCTCTTATGGGCGGC 4133  
 DB 4051 ATCTGGCAAAAGATTCACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGTGA 4110  
 QY 4134 TTTGACTCAAGAACCGCCTCTCAGATCCTCATCAAAACACAGCCTGTCTCTGCGAAT 4193  
 DB 4111 TTCGACTTAAACACCTCTCTCCACAGATTCATCAAGAACACCCCGGTACTGCGAAT 4170  
 QY 4194 CCTCCGCGGAGTTTCACTACAAAGTTTGTCTCATTCATCAACCAATACCTCCACAGGA 4253  
 DB 4171 CCTTCGACCACTTCACTGCGCGCAAGTTTGTCTCTCATCAACACAGTACTCCACGGA 4230  
 QY 4254 CA-AGTGTGTGAATTTGAATGGAGCTGCGAAGAAAGAAACAGCAGCGCTGGAATCC 4312  
 DB 4231 CACGCTCAGCGTGAAGTGAAGTGGAGCTGCGAAGAAAGAAACAGCAGCGCTGGAATCC 4290  
 QY 4313 CGAAGTGCATACATCCAAATTTATGCAAAATCTGCCAAGTGTATTTTACTGTGACAA 4372  
 DB 4291 CGAAATTCAGTACCTTCCAACTACAAACAGTGTGTATCTGTGACTTACCGTGATAC 4350  
 QY 4373 CAATGACTTTATACGCTCGCCCATTTGGCAGCCGTTACCTTACCGTCCCTGTA 4432  
 DB 4351 TAATGGCGGTATTCAGAGCCTCGCCCATTTGGCAGCAGATACCTGACTGTATCTGTA 4410

QY 4433 ATTACGTGTTAATCAATAAACCGGTGATTCGTTTCACTGTAACCTTGTCTCTCTGCT 4492  
 DB 4411 ATTGCTGTTAATCAATAAACCGGTGATTCGTTTCACTGTAACCTTGTCTCTCTGCT 4470  
 QY 4493 TCTTATCTTATCGGTACCATGCTTATAGCTTACATTAACCTGTTGCTGCTGCT 4552  
 DB 4471 TCTTATCTTATCGGTACCATGCTTATAGCTTACATTAACCTGTTGCTGCTGCT 4530  
 QY 4553 GATTAAGACTTACGTATCGGTTTACCCCTAGTGAAGTGTGCGCACTCCCTCTGCT 4612  
 DB 4531 CTACAGGA-----ACCCCTAGTGAAGTGTGCGCACTCCCTCTGCT 4574  
 QY 4613 GCGTTCGCTCGCTCGGTGGGCTGCGGACCAAGTCCGACAGACGCGAGAGCTGCTC 4672  
 DB 4575 GCGTTCGCTCGCTCGGTGGGCTGCGGACCAAGTGTGCGCACTCCCTCTGCT 4634  
 QY 4673 TCCCGGCCCCACCGAGCGGAGCGCGGACAGAGAGGAGTGGCAA 4718  
 DB 4635 GGGCGGCTCAGTGAGCGAGCGGAGCGCGGACAGAGAGGAGTGGCAA 4680

RESULT 8  
 AAH41481  
 ID AAH41481 standard; DNA; 4675 BP.  
 XX  
 AC AAH41481;  
 XX  
 DT 23-AUG-2001 (first entry)  
 XX  
 DE Adeno-associated virus (AAV) plasmid pAV1 Avari nucleotide sequence.  
 XX  
 KW Adeno-associated virus; AAV; gene transfer; adenovirus; chromosome 19;  
 KW medicine; cell engineering; genetic engineering; treatment;  
 KW embryological engineering; ds.  
 OS Adeno-associated virus.  
 XX  
 PN WO200132899-A1.  
 PD 10-MAY-2001.  
 XX  
 PF 23-OCT-2000; 2000WO-JP07373.  
 XX  
 PR 29-OCT-1999; 99JP-0308839.  
 XX  
 PA (TAKI ) TAKARA SHUZO CO LTD.  
 XX  
 PI Ueno T, Matsumura H, Tanaka K, Iwasaki T, Ueno M, Fujinaga K;  
 PI Asada K, Kato I;  
 XX  
 DR WPI; 2001-316450/33.  
 XX  
 PT Gene transfer method for highly efficient introduction of foreign genes  
 PT to cells in humans particularly by integration specifically into AAVS1  
 PT site of chromosome 19 for transformation, applicable e.g. in disease  
 PT treatment -  
 XX  
 PS Example 1; Page 28-31; 38pp; Japanese.  
 XX  
 CC The present invention describes a gene transfer method for highly  
 CC efficient introduction of foreign genes to cells in humans particularly  
 CC by integration specifically into AAVS1 site of chromosome 19 for  
 CC transformation, applicable e.g. in disease treatment. The method  
 CC comprises transferring into the cells (with the use of an adenovirus  
 CC vector) a nucleic acid which has a sequence provided with adeno-  
 CC associated virus (AAV)-originated inverted terminal repeats (ITRs) in  
 CC both sides of the target foreign gene to be transferred, a second nucleic  
 CC acid which has an AAV-originated rep gene and a promoter for expressing  
 CC this gene and carries a stuffer sequence inserted into it sandwiched in  
 CC two recombinase-recognition sequences and located between the rep gene  
 CC and promoter and expressing the Rep protein under the action of the  
 CC recombinase in the cells obtained previously to integrate the target  
 CC foreign gene into the chromosomal DNA. The method is for transferring a

CC foreign gene into cells particularly in human, especially by integration  
CC specifically into AAVS1 site of chromosome 19 for transformation, which  
CC is applicable in medicine, as well as cell, genetic and embryological  
CC engineering e.g. in disease treatment. The method is highly efficient,  
CC with use of adenovirus vectors, nucleic acids and other sequences  
CC including rep genes and promoters, by expressing rep proteins to  
CC integrate target foreign gene into chromosomal DNA through action of  
CC recombinase. The present sequence represents an AAV plasmid pAV1 Avail  
CC nucleotide sequence which is used in an example from the present  
CC invention.

XX Sequence 4675 BP; 1198 A; 1262 C; 1251 G; 964 T; 0 other;

Query Match 64.8%; Score 3055.8; DB 22; Length 4675;  
Best Local Similarity 79.9%; Pred. No. 0;  
Matches 3779; Conservative 0; Mismatches 887; Indels 61; Gaps 13;

QY 1 TTGCCCACTCCCTCTCTGCGCGCTCGCTCGGTGGGGCTCGGACCAAGGTCCGC 60  
DB 1 TTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGCGCAACCAAGTCCGC 60  
QY 61 AGACGGCAGAGCTCTGCTCTGCGCGCGCGCGCGCGAGCGGCGGCGAGAGAGGAGTG 120  
DB 61 CGACGCCCGGCGCTTGGCCGGCGGCTCACTGAGCGAGCGGCGGCGAGAGAGGAGTG 120  
QY 121 GGCAACTCCATCACTAGGGGTAATCGCGAAGCGCTCCACGCTGCCGCTCAGCGCTGA 180  
DB 121 GCCAACTCATCACTAGGGGT-----TCTGAGAGGGGTGAGTCTGTGA 163  
QY 181 CGTAATTAAGTCATAGGG--GAGTGTCTGTATTAAGCTGTCACTGAGTGTCTTTGC 237  
DB 164 CGTGAATTAAGTCATAGGGTGAAGGAGGCTGTATTAAGAGTCACTGAGTG-TTTTGC 222  
QY 238 GACATTTTGGCAGACCAACAGTGGCCATTAGGGTATATATGCGCGAGTGAAGCGAGAGAT 297  
DB 223 GACATTTTGGCAGACCAATGTGTACGCTGGGTATTAAAGCCCGAGTGAACGACGAGGT 282  
QY 298 CTCCATTTTG-AACCGGAAATTGAACGAGCAGCAGCCATGCCGGCTTCTACGAGATCG 356  
DB 283 CTCCATTTTGAAGCGGAGGTTTGAACGCGCAGCCGATGCCGGGTTTACGAGATTG 342  
QY 357 TGATCAAGGTCCGAGCGACCTGAGCAGCACTGCCGGCATTTCTGACTGTTTGA 416  
DB 343 TGATTAAGTCCCGACGACCTTGACGGGATCTGCCGCGCATTTCTGACAGCTTTTGA 402  
QY 417 GCTGGTGGCGGAGAGGATGAGAGTGGCCCGGATTTCTGACATGATCTGAATCTGA 476  
DB 403 ACTGGGTGGCGAGAGGATGGAGTGGCGGCGAGATTCTGACATGATCTGAATCTGA 462  
QY 477 TTGAGCAGGACCCCTGACCGTGGCGGAGAGCTGCAGCGGACTTCTGTCCAATGGC 536  
DB 463 TTGAGCAGGACCCCTGACCGTGGCGGAGAGCTGCAGCGGACTTCTGACGGAATGGC 522  
QY 537 GCCGCGTGAAGGCCCCCGGAGGCGCTTCTTTGTTCAAGTTCAAGAGGGCGAGTCTT 596  
DB 523 GCCGTGAGTGAAGGCCCGCGAGGCCCTTTTCTTGTGCAATTGAGAGGAGAGAGCT 582  
QY 597 ACTTCCACTCCATATTTCTGTGAGAGCAGCGGGGTCAATCCATGCTGTGGCGCT 656  
DB 583 ACTTCCACTCACTGCTCGTGAACCAACCGGGGTGAATCCATGCTTTTGGGAGCTT 642  
QY 657 TCCTGAGTCAATTAGGAGCAAGCTGTGAGACCATCTACCGCGGATCGAGCGGACCC 716  
DB 643 TCCTGAGTCAATTGCGGAAAACTGATCAGAGATTACCGCGGATCGAGCGGACTT 702  
QY 717 TGCCCAACTGTTGCGGCTGACCAAGCGGTAATGGCGCGGAGGGGGAACAAGTGG 776  
DB 703 TGCCAAACTGTTGCGGCTCACAAGACCAAGAAATGGCGCGGAGGGGGAACAAGTGG 762  
QY 777 TGAGCAGTGTCTATCCCAACTACTCTGCTGCCAAGACTCAGCCGAGCTGCAAGTGG 836  
DB 763 TGAGTAGTGTCTATCCCAATTACTTGTCTCCCAAAACCCAGCTGAGCTCAAGTGG 822

QY 837 CGTGACTAATAGAGAGTATATAAGCGCTGTGTAACCTGGCCGAGCGCAACGGC 896  
DB 823 CGTGACTAATATGAAACAGTATTTAAGCGCTGTGTAATCTCAGGAGCGTAAACGT 882  
QY 897 TCGGCGGAGCAGCAGCTGACCGACGTGACGAGCAGACCGAGAGAGCAAGAGAACTGA 956  
DB 883 TGGTGGCGCAGCATCTGACGACGTGTCGAGACGAGAGAGAGCAAGAAAGATCAGA 942  
QY 957 ACCCAATTCTGACGCGCTGTCTATCCGGTCAAAAACCTCCGCGCTACATGAGCTGG 1016  
DB 943 ATCCCAATTCTGATGCGCGGTGATCAGATCAAAAACCTTCAAGGATCATGAGCTGG 1002  
QY 1017 TCGGGTGGCTGTGAGACCGGGGATCACCCTCCGAGAGAGTGAATCCAGAGAGCAGG 1076  
DB 1003 TCGGGTGGCTGTGAGACAGGGGATTAACCTCGAGAGAGTGAATCCAGAGAGCAGG 1062  
QY 1077 CCTGTACATCTCTTCAACGCGCTTCCAACTCGCGTCCAGATCAAGGCGCTCTGG 1136  
DB 1063 CCTATACATCTCTTCAATGCGGCTCCAACTCGCGTCCCAATCAAGGCTGCTGG 1122  
QY 1137 ACAATGCCGCAAGATCAATGCGCTGACCAAAATCCGCGCGGACTACTGTAGGCCCG 1196  
DB 1123 ACAATGCCGGAAGATTAATGAGCTGACTAAACCGCGCGGACTACTGTAGGCCAGC 1182  
QY 1197 CTCCGCGCGGAGCATTAACCAACCGCATCTACCGCATCTGAGCTGAACGGCTACG 1256  
DB 1183 AGCCCGTGAAGACATTTCCAGCAATCGGATTTATAAATTTTGAACATAACGGGTACG 1242  
QY 1257 AACCTGCTACGCGCGCTCCGCTTTCTCGGCTGGGCGCCAGAAAAAGTTCCGGAAGCGCA 1316  
DB 1243 ATCCCAATATGCGGCTTCCGCTTTCTGGATGGGCCAGAAAAAGTTCCGCAAGAGA 1302  
QY 1317 ACAACATCTGGCTGTTGGCGGCGCACCAACGCGGCAACCAATCGCGGAAGCATCG 1376  
DB 1303 ACAACATCTGGCTGTTGGCGCTGCAACTACCGGGAAGCAACATCGCGAGGCCATAG 1362  
QY 1377 CCCAGCGCTGCCCTTCTACGGCTGCGTCAACTGAGCAATGAGAACTTCCCTCAATG 1436  
DB 1363 CCCAGCTGTGCCCTTCTACGGGTGCGTAACTGAGCAATGAGAACTTCCCTCAACG 1422  
QY 1437 ATTGCTCGACAGATGATCTGTGGAGAGGCGAAGATGACGCGCAAGGTCTGG 1496  
DB 1423 ACTGTGCGACAGATGATGATCTGTGGAGAGGCGAAGATGACCGCAAGGTCTGG 1482  
QY 1497 AGTCCGCCAAGGCAATTTCTGCGCGGCGAGCAAGGTGCGCGTGAACCAAGTGCAGTGT 1556  
DB 1483 AGTCCGCCAAGCAATTTCTGAGAGAGCAAGGTGCGCGTGAACCAAGTGCAGTGT 1542  
QY 1557 CCGCCAGATCGACCCCGGCGGATGCTGATCTGATCTCAACCAACATGTGCGCGTGA 1616  
DB 1543 CCGCCAGATGAGCCCGGCGGATGCTGATCTGATCTCAACCAACATGTGCGCGTGA 1602  
QY 1617 TTGACGGGAACAGCAGCACTTTCAGACAGCAGCGGCTGAGAGCGGATGTTCAAT 1676  
DB 1603 TTGACGGGAACCTCAACGACCTTTCAGACAGCAGCGGCTGAGAGCGGATGTTCAAT 1662  
QY 1677 TTGAATCAACCGCGCTGAGCATGACTTTGGCAAGGTGAACAAGCAGAGTCAAG 1736  
DB 1663 TTGAATCAACCGCGCTGAGCATGACTTTGGCAAGGTGAACAAGCAGAGTCAAG 1722  
QY 1737 AGTTCTTCGCTGGGCGGAGATCACTGACCGAGGTGCGCATGAGTTCTACGTCAGAA 1796  
DB 1723 AGTTCTTCGCTGGGCGGAGATCACTGAGGTGAGGTGAGCATGAATTTCTACGTCAGAA 1782  
QY 1797 AGGTGGAAGCAACAAAGACCGCGCGGATGAGCGGATGAAGGAGAGCGGCGGAGCGG 1856  
DB 1783 AGGTGGAAGCAACAAAGACCGCGCGGATGAGCGGATGAAGGAGAGCGGCGGAGCGG 1842  
QY 1857 CCTGCCCTCAGTGGCGGATCATGAGCTCAGACGCGGAGAGAGCTCCGTTGACTTTG 1916  
DB 1843 TGGCGAGTCAAGTGGCGGATCATGAGCTCAGACGCGGAG--AGCTTCAGTCAACTACG 1899  
QY 1917 CCGACAGGTACCAAAAACAAATGTTCTCTGTCACGCGGCGATGCTTCAAGTGTGTTCCCT 1976



1900 CAGACAGGTACCAAAACAATGTTCTCGTCACGTGGGCATGAATCTGATGCTGTTCCCT 1959  
QY 1977 GCAAGACATCGAGAGATGAATCAGAATTTCACATTGCTTCACGCGGACGAGAG 2036  
Db 1960 GCAGACAATGCGAGAGATGAATCAGAATTAATCTGCTTCACTCAGGACAGAAAG 2019  
QY 2037 ACTGTTCAAGTGTCTCCCGCGGTGTGAGAAATCTCAACCGGT---CGTCAGAAAGAGA 2093  
Db 2020 ACTGTTAGAGTGTCTCC---CGTCAGAAATCTCAACCCGTTCTGTGTCACAAAAGG 2076  
QY 2094 CGTATCGAAACTCTGTGCAATTCATCATCTGTGGGGCGGCTCCCGAGATTGCTGCT 2153  
Db 2077 CGTATCAGAACTGTGTACATTCATCATATGAGGAAAGTGCCAGA---CGCTTGCA 2133  
QY 2154 CGGCTGCGATCTGTGTAACGTGGAACCTGGAATCTGTGTTCTGAGCAATAATGACTT 2213  
Db 2134 CTGCTGCGATCTGTGTAATGTGATTGATTTGATGACTGCATCTTGAACATAAATGATTT 2193  
QY 2214 AAACCAAGTATGCTGTCCCGATGTTATCTTCCAGATTGGCTCGAGGACAACTCTGAG 2273  
Db 2194 AAATCAGTATGCTGTCCGATGTTATCTTCCAGATTGGCTCGAGGACACTCTCTGAA 2253  
QY 2274 GGCATTGCGAGTGTGGGACTTGAAACCTTGAGCCCGAAGCCCAAGCCAAACGAGCAA 2333  
Db 2254 GGAATTAAGACAGTGTGGAAAGCTCAAACTTGCCCAACCAACCAAGCCGAGAGCGG 2313  
QY 2334 AAGCAGACGACGCGCGGGTCTGTGCTTCTGCTCAAGTACCTCGAACCTTCAAC 2393  
Db 2314 CATAGGACGACGAGGGGTCTGTGCTTCTGTGGTACAAATACCTCGGACCTTCAAC 2373  
QY 2394 GGAATCGACAAAGGGGAGCCCGTCAACGCGCGGACGAGCGCCCTGAGCAAG-ACAA 2452  
Db 2374 GGAATCGACAAAGGGGAGCCCGTCAACGAGGACGAGCGCGCCCTGAGCAAGTACAA 2433  
QY 2453 GGCCTACGACGACGCTCAAGCGGGTGAACAATCCGTAACCTGCGTATTAACAAGCCGA 2512  
Db 2434 AGCCTACGACCGGCGAGCTCGACAGCGGAGACAAACCTGTAAGTACAAACGCGCGA 2493  
QY 2513 CGCCGAGTTTCAGAGAGCGTCTGCAAGAAATACGTTTGTGGGCAACCTCGGGCGAGC 2572  
Db 2494 CGCGGAGTTTCAGAGAGCGCTTAAAGAAATACGTTTGTGGGCAACCTCGGACGAGC 2553  
QY 2573 AGTCTTCAGGCGCAAGAGCGGGTCTCGAACTCTCGGTCTGTTGAGGAAAGCGCTAA 2632  
Db 2554 AGTCTTCAGGCGCAAGAGGGTCTTGAACCTCTGGGCTGTGAGGAACTGTGTA 2613  
QY 2633 GACGCTCCTGGAAGAAAGTCCGCTAGAGCACTGCGCACAAAGAGCCGAGACTCTCTC 2692  
Db 2614 GACGCTCCGGAAGAAAGAGCGCGTAGAGCACTCTGTGAGCCGAGACTCTCTC 2673  
QY 2693 GGGCATCGCAAGAGAGCGGCGGCTTAAAGAGAGCTCAATTTGCTCAGACTGG 2752  
Db 2674 GGGAAACCGGAAGCGGGCCAGAGCCTGCAAGAAAGATTTGATTTGCTCAGACTGG 2733  
QY 2753 CGACTCAGAGTCACTCCCGATCCACAACTCTCGGAGAACTCCAGCAACCCCGCTGC 2812  
Db 2734 AGACGAGACTCAGTACTGACCCCGAGCTCTCGGACAGCCAGCAGCCTCTG 2793  
QY 2813 TGTGGACCTTACTCAATGCTTCAAGCGGTGGCGCAACCAATGGCAGACAAATAAGAGG 2872  
Db 2794 TCTGGAACTAATACATGCTTACAGGCACTGGCGCAACCAATGGCAGACAAATAAGAGG 2853  
QY 2873 CGCCGACGAGTGGTAAATGCTCAGGAATTTGCAATTTGCAATGCTGCGGCGA 2932  
Db 2854 CGCCGACGAGTGGTAAATTTCTCCGGAATTTGCAATTTGCAATGCTGCGGCGA 2913  
QY 2933 CAGAGTCAATCAACCAAGCAACCGGAGCTGCGGCTTGCACCTACAAATAACCACTCTA 2992  
Db 2914 CAGAGTCAATCAACCAAGCAACCGGAGCTGCGGCTTGCACCTACAAATAACCACTCTA 2973  
QY 2993 CAAGCAATCTCAGTCTTCAACGGGGCGCAGCAACGACCACTACTTGGCTACAG 3052

Db 2974 CAAACAATTTCCAGCCAATCA---GGAGCCTCGAAGCAATCACTACTTTGGCTACAG 3030  
QY 3053 CACCCCTGGGGTATTTTGAATTTCAACAGATTTCACTGCCACTTTTCAACCAAGTACTG 3112  
Db 3031 CACCCCTGGGGTATTTTGAATTTCAACAGATTTCACTGCCACTTTTCAACCAAGTACTG 3090  
QY 3113 GCAGCACTCATCAACAACAATTTGGGATTTCCGGCCCAAGAGACTCAACTTCAACTCTT 3172  
Db 3091 GCAAGACTCATCAACAACAAGTGGGATTTCCGACCCAGAGACTCAACTTCAAGCTCTT 3150  
QY 3173 CAACATCAAGTCAAGAGGTCAGACGAATGATGGCGTCAAAACCATGCTAATAACT 3232  
Db 3151 TAACATTCAGTCAAGAGGTCAGACGAATGACGGTACGACGATTTGCCAATAACT 3210  
QY 3233 TACAGCAGCGTTCAAGTCTTTCGGACTCGGAGTACGACTTCCGTACGCTCGGCTC 3292  
Db 3211 TACAGCAGCGTTCAAGTCTTTCGGACTCGGAGTACGACTTCCGTACGCTCGGCTC 3270  
QY 3293 TGCGCAGGCGTGCCTCCCTCGTTCGCGGCGGACGCTGTATGATTTCCGCAATACG 3352  
Db 3271 GCGCAGTCAAGATGCTCCCGCGTTCAGAGCAGAGCTTCAATGTCACAGATATG 3330  
QY 3353 CTACCTGACGCTCAACAATGGCAGCCAGCCGTGGAGCTTCACTTTTACTGCTGGA 3412  
Db 3331 ATACCTCAACCTGAACAACGGGAGTCAAGCAGTAGAGCGCTTCAATTTACTGCTGGA 3390  
QY 3413 ATATTTCCCTTCTCAGATGCTGAGAACGGGCAACAATTACCTTACGCTACACTTTGA 3472  
Db 3391 GTACTTCTTCTCAGATGCTGCTACCGGAACAATTACCTTACGCTACACTTTGA 3450  
QY 3473 GGAAGTGCCTTTCACAGCAGCTACGCGCACAGCCAGAGCTGGAACCGGTGATGAATCC 3532  
Db 3451 GGAAGTGCCTTTCACAGCAGCTACGCTCAACGAGAGTCTGGAACCGGTGATGAATCC 3510  
QY 3533 TCTCATGACCAATATCTGTATTACTGAAACAGAACTCAAAATCAGTCCGGAATGCCA 3592  
Db 3511 TCTCATGACCAATATCTGTATTACTGAAACAGAACTCAAAATGGAACCACTCAC 3570  
QY 3593 AAACAAGACTTGTGTTTACCGGTGGGTCTCCAGCTGCAATGTTTACGCCCAAAA 3652  
Db 3571 GCAGTCAAGCTTCAATTTCTCAGAGCGGAGAGTGAATTTGGGACAGTCTAGAA 3630  
QY 3653 CTGGTACCTGAGCCCTGTATCGGACAGAGCGGCTTTTAAACCAAAAACAGACAA 3712  
Db 3631 CTGGTCTCTGAGCCCTGTATCGGACAGAGAGTATCAAGAATCTGCGGATTAACA 3690  
QY 3713 CAACAGCAATTTTACCTGAGCTGTGCTTCAAAATATACTCAATGGCGGTGAATCAT 3772  
Db 3691 CAACAGTGAATCTCGTGAATGAGCTTCAAGTACCACTCAATGGCAGAGACTCTCT 3750  
QY 3773 CATCAACCTGAGCACTGCTATGGCTTCAACAAGACGAGCAAGCAAGTCTTTCCAT 3832  
Db 3751 GGTGAATC-----CGGCATGGCAAGCCACAAGAGCATGAAGAAAGTTTCTCTCA 3804  
QY 3833 GAGCGGTGTCATGATTTTGAAGAAAGAGCGCGCGGAGCTTCAACAATGCAATTTGCA 3892  
Db 3805 GAGCGGTGTCATGATTTTGAAGCAAGGCTCAGAGAAAACAATGTGAACATTTGAAA 3864  
QY 3893 TGTCAATTAAGAGAGAAATTAAGCCACTAACCTGTGGCCACCGAAAGATT 3952  
Db 3865 GGTCAATTAAGAGAGAAATCGAACAACCAATCCGTGGCTACGAGAGATA 3924  
QY 3953 TGGACCGTGGCAGTCAATTTCCAGAGCAGACAGACCTGCGACCGAGATGTGA 4012  
Db 3925 TGGTCTGTATTAACCAACCTCAGAGAGCAAGCAAGCAAGTACCGAGATGTGA 3984  
QY 4013 TGTATGGAAGCATTAACCTGGAATGCTGTGGCAAGATAGACGTTACTGCAAGGTTCC 4072  
Db 3985 CACACAGGCGTCTTCCAGGAGTGTGTCAGAGACAGAGATGTACTTCAAGGCGC 4044  
QY 4073 CATTTGGGCCAAATTTCTCAGACAGATGACACTTTACCCGTCTCTTATGGGCGG 4132  
Db 4045 CATTTGGGCCAAATTTCCACAGCGGAGCATTTTCAACCCCTCTCCCTCATGGGTGG 4104



QY 4133 CTTTGGACTCAAGAACCCGCTCCTCAGATCCTCATCAAAAACACGCTGTTCCTGCGAA 4192  
Db 4105 ATTCGGAETTAACACACCCTCCTCCACAGATTCATCAAGAACACCCCGGTACTGCGAA 4164  
QY 4193 TCCTCCGCGGAGTTTTCAGCTACAAAGTTTGCTTCATTCATCAACCAATACTCCACAGG 4252  
Db 4165 TCCTTCGACCACCTTCAGTGGCGAAAGTTTGCTTCCTTCATCAACAGTACTCCACAGG 4224  
QY 4253 ACA-AGTAGAGTGTGAAATTGAATGGAGCTGCAGAAAGAAAACAGCAAGCGCTGGAATC 4311  
Db 4225 ACACGGTCACGCTGAGATCGAGTGGAGCTGCAGAAAGAAAACAGCAACGCTGGAATC 4284  
QY 4312 CCGAAGTCAGTACACATCCAAATTAAGCAAAATCTGCCAAGCTGATTTTACTGTGACA 4371  
Db 4285 CCGAATTCAGTACACTCCAACTACAACAAGTCTGTTAATCGTGAAGTACCGTGAATA 4344  
QY 4372 ACAATGACTTTATATCTAGAGCTGCGCCCATTTGGCACCCGTTACCTTACCCGTCCTGT 4431  
Db 4345 CTATGCGGTGATTCAGAGCCCTGCGCCCATTTGGCACCAAGATACCTGACTCGTATCTGT 4404  
QY 4432 AATTACGTGTTAATCAATAAACCGGTTGATTCGTTCAAGTTGAATTGGTCTCCTGTC 4491  
Db 4405 AATTGCTGTTAATCAATAAACCGGTTAATTCGTTCAAGTTGAATTGGTCTCCTGTC 4464  
QY 4492 TTCTTATCTTATCGGTTACCAAGTTTATAGCTTACACATTAAGTCTGTGCTGCTG 4551  
Db 4465 TTCTTATCTTATCGGTTACCAAGTTTATAGCTTACACATTAAGTCTGTGCTGCTG 4524  
QY 4552 CGATAAAGACTTACGTCATCGGTTAACCCCTAGTAGAGAGTTGCCCACTCCCTCTCTG 4611  
Db 4525 ACTACAAGGA-----ACCCCTAGTAGAGAGTTGCCCACTCCCTCTCTG 4568  
QY 4612 CGCGCTCGCTCGCTCGGAGGCGCGGAGCAAAAGTCCGAGACGCGAGAGCTCTGCT 4671  
Db 4569 CGCGCTCGCTCGCTCGGAGGCGCGGAGCAAAAGTCCGAGACGCGAGAGCTCTGCT 4628  
QY 4672 CTGCGGCGGCGGCGGAGCGGAGCGGCGGAGAGAGGAGTGGGCAA 4718  
Db 4629 CGGCGGCGGCGGAGCGGAGCGGCGGAGAGAGGAGTGGGCAA 4675

RESULT 9  
ID AAF89931 standard; DNA; 4675 BP.  
XX AAF89931;  
AC AAF89931;  
DT 06-AUG-2001 (first entry)  
XX Nucleotide sequence of an adeno-associated virus Rep78 protein.  
DE adeno-associated virus; AAV; Rep78; papillomavirus-associated disease;  
KW cancer; HIV-associated disease; ss.  
XX Adeno associated virus.  
OS  
FH Key Location/Qualifiers  
FT CDS 321..2186  
FT FT /\*tag= a  
FT FT /product= "Rep78"  
XX PN WO200132711-A2.  
XX PD 10-MAY-2001.  
XX PF 23-OCT-2000; 2000WO-US29187.  
XX PR 21-OCT-1999; 99US-0160608.  
XX PA (UYAR-) UNIV ARKANSAS.  
XX PI Hermonat PL;

XX WPI; 2001-343480/36.  
DR P-PSDB; AAB83954.  
XX  
PT New adeno-associated virus AAV Rep78 mutant, useful for treating  
PT papillomavirus-associated diseases, cancer, or human immunodeficiency  
PT virus-associated diseases  
PS  
XX Disclosure; Fig 15; 53pp; English.  
CC The present sequence encodes an adeno-associated virus (AAV) Rep78  
CC protein. The specification describes an adeno-associated virus (AAV)  
CC Rep78 mutant, comprising an AAV Rep78 modified protein that possesses  
CC different biochemical and biological functions as compared to the  
CC wild-type AAV Rep78 protein. The AAV Rep78 mutant and associated DNA  
CC sequences can be used in a pharmaceutical composition. The pharmaceutical  
CC composition can be used for treating papillomavirus-associated diseases,  
CC cancer, or HIV-associated diseases.  
XX  
SQ Sequence 4675 BP; 1198 A; 1262 C; 1251 G; 964 T; 0 other;

Query Match 64.8%; Score 3055.8; DB 22; Length 4675;  
Best Local Similarity 79.9%; Pred. No. 0;  
Matches 3779; Conservative 0; Mismatches 887; Indels 61; Gaps 13;

QY 1 TTGCCCCCTCCCTCTCTGCGCGCTCGCTCGCTCGGCGGCGCTCGGACCAAGTCCGC 60  
Db 1 TTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCGCTCGGCGGCGGCGGACCAAGTCCGC 60  
QY 61 AGACGGCAGAGCTCTGCTCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120  
Db 61 CGACGCCCGGCTTGTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120  
QY 121 GCGAATCCATCACTAGGCGGTAATCGGAAGCGGCTCCCAAGCTGCCGCTGAGCGGCTGA 180  
Db 121 GCGAATCCATCACTAGGCGGTAATCGGAAGCGGCTCCCAAGCTGCCGCTGAGCGGCTGA 163  
QY 181 CGTAATTTAGCTATAGG--GAGTGTCTGTATTAAGTGTACGTGAGTGTCTTTTGC 237  
Db 164 CGTGAATTTAGCTATAGGTTAAGGAGTCTGTATTAAGGTCACGTGATG-TTTTGC 222  
QY 238 GACATTTTGGCAGACCACTGCGGCTTTAGGGTATATAGCGGAGTGAAGCGAGGAT 297  
Db 223 GACATTTTGGCAGACCACTGCTGCGGCTTTAGGGTATATAGCGGAGTGAAGCGAGGAT 282  
QY 298 CTCCATTTTG-ACCGGAAATTTGAACGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 356  
Db 283 CTCCATTTTGAAGCGGAGGTTTGAACGCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 342  
QY 357 TGATCAAGGTGCGGAGCGGCTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 416  
Db 343 TGATTAAGTCCCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 402  
QY 417 GCTGGGTGCGGAGAGGAATGGGAGCTGCGGCGGAGTTTGAACATGATGATGATGATGATG 476  
Db 403 ACTGGGTGCGGAGAGGAATGGGAGCTGCGGCGGAGTTTGAACATGATGATGATGATGATG 462  
QY 477 TTGAGCAGGCAACCCCTGACCGGTGCGGAGAGCTGAGCGGAGCTTCTGCTGCAATGCG 536  
Db 463 TTGAGCAGGCAACCCCTGACCGGTGCGGAGAGCTGAGCGGAGCTTCTGCTGCAATGCG 522  
QY 537 GCGCGGTGAGTAAGGCGCGGAGGCGGCTTCTGTTGTTAGTTGAGTAAGGCGGAGTCT 596  
Db 523 GCGGTGAGTAAGGCGCGGAGGCGGCTTCTGTTGTTAGTAAGGCGGAGTCTGAGTAAGT 582  
QY 597 ACTTCCACCTCCATATTTCTGTGAGAGCAGCGGCTCAATCCATGCTGCTGCGGCGGCT 656  
Db 583 ACTTCCACATGACGTGCTGTGGAACACCGGCTGAATCCATGCTGCTGCGGCGGCT 642  
QY 657 TCCTGAGTCAGATTAGGACAAGCTGTGAGACCACTACCGCGGAGTGAAGCGGAGCC 716  
Db 643 TCCTGAGTCAGATTGCGGAAAAAAGTATTGAGAAATTAACCGCGGAGTGAAGCGGAGCT 702

QY 717 TGCCCACTGCTTCCGGGTGACCAAGACGGTATGCGCCGAGGGGGAACAAGTGG 776  
 |||||  
 Db 703 TGCCAAACTGCTTCCGGGTACAAAGACCAAGATGGCGCCGAGGGGGAACAAGTGG 762  
 QY 777 TGGACGAGTGTACATCCCACTACCTCTGCCCCAAGACTCAGCCGAGCTGAGTGG 836  
 |||||  
 Db 763 TGGATGAGTGTACATCCCACTACCTGCTCCCCAAAACCCAGCTGAGCTCCAGTGG 822  
 QY 837 CGTGACTACATGAGAGGATATATAGCGCTGTTGAACTTGAGCCGCAAGCCGCAACGGC 896  
 |||||  
 Db 823 CGTGACTACATGAGAACAGTATTTAAGCCCTGTTGAATCTCAGCGAGCGTTAAACGGT 882  
 QY 897 TCGTGGCGGAGCACTGACCCAGCTCAGCCAGACCCAGAGCAGAACAGAGAACTTGA 956  
 |||||  
 Db 883 TGGTGGCGGAGCATCTGACGCACTGTGCGAGACGAGAGCAGAACAGAGAAATCAGA 942  
 QY 957 ACCCCAATTCTGACGCGCTGTCTATCCGCTCAAAAACTCCGCGCGCTACATGAGCTGG 1016  
 |||||  
 Db 943 ATCCCAATTCTGATGCGCCGCTGATCAGATCAAAAACTTCAAGCAGTACATGAGCTGG 1002  
 QY 1017 TCGGCTGCTGCTGAGACCGGGCATCACTCCGAGAGCAGTGTGATCCAGAGACAGG 1076  
 |||||  
 Db 1003 TCGGCTGCTGCTGAGACAGGGGATTAACCTCGAGAGAGAGTGTGATCCAGAGACAGG 1062  
 QY 1077 CCTCGTACATCTCTTCAACGCGCTTCAACTCGCGCTCCAGATCAAGCGCTGG 1136  
 |||||  
 Db 1063 CCTCATACATCTCTTCAATGCGGCTCCAACTCGCGCTCCAAATCAAGCTGCTGG 1122  
 QY 1137 ACAATGCGGCAAGATCATGCGCTGACCAAAATCCGCGCGCACTACCTGTAGGCCCCG 1196  
 |||||  
 Db 1123 ACAATGCGGCAAGATTAATGAGCTGATTAACCGCCCCGCACTACCTGTGGGCCAGC 1182  
 QY 1197 CTCCGCGCGGAGCATTAACCAACCGCATCTACCGCATCTGAGCTGAACGGCTAGC 1256  
 |||||  
 Db 1183 AGCCCGTGGAGCATTTCCAGCAATCGATTTTAAATTTGAACTTAAACGGGTACG 1242  
 QY 1257 AACCTGCTACGCGCGCTCTCTTCTCGCTGGGCCAGAAAAGTTGCGGAAGCGCA 1316  
 |||||  
 Db 1243 ATCCCAATTGCGGCTTCCGCTCTTCTGGAGTGGGCCACGAAAAGTTGCGCAAGAGA 1302  
 QY 1317 ACACCATCTGCTGTTGGGCCGAGCAACGAGCAAGACCAACATCGCGGAAGCCATCG 1376  
 |||||  
 Db 1303 ACACCATCTGCTGTTGGGCCGAGCACTACCGGGAAGACCAACATCGCGGAAGCCATAG 1362  
 QY 1377 CCCACGCGCTTCTTACGCGCTGCTCACTGGAACCAATGAACTTCCCTTCAATG 1436  
 |||||  
 Db 1363 CCCACATGCTGCTTCTACGGGTGCGTAACCTGGAACCAATGAACTTCCCTTCAACG 1422  
 QY 1437 ATTGCTGCAAGATGCTGATCTGCTGGAGAGAGGCAAGTGAACGCGCAAGTCTGG 1496  
 |||||  
 Db 1423 ACTGTGTCGAAGATGCTGATCTGCTGGAGAGAGGGAAGTGAACGCGCAAGTCTGG 1482  
 QY 1497 AGTCCGCAAGGCTATCTCGGCGGAGCAAGGTGCGGCTGGAACCAAAAGTGAAGTCT 1556  
 |||||  
 Db 1483 AGTCCGCAAGGCTATCTCGGAGGAAGCAAGGTGCGGCTGGAACCAAAATGCAAGTCT 1542  
 QY 1557 CCGCCAGATGACCCCAACCCCGTATCTCACTCAACCAACCAACATGTGCGCGTGA 1616  
 |||||  
 Db 1543 CCGCCAGATGACCCCACTCCCGTATCTCACTCAACCAACCAACATGTGCGCGTGA 1602  
 QY 1617 TTGACGGAAGCAACCACTTGCAGACCAAGCCGTTGACAGACCGGATGTTCAAT 1676  
 |||||  
 Db 1603 TTGACGGAAGCAACCACTTGCAGACCAAGCCGTTGACAGACCGGATGTTCAAT 1662  
 QY 1677 TTGACCTACCCGCTGTGAGCATGCTTTGGCAAGGTGACAAAGCAGAGTCAAG 1736  
 |||||  
 Db 1663 TTGACCTACCCGCTGTGAGCATGCTTTGGCAAGGTGACAAAGCAGAGTCAAG 1722  
 QY 1737 AGTCTTCCGCTGGCGCAGGATCAAGTGAACGAGTGGCGCATGAGTCTTCAAGAA 1796  
 |||||  
 Db 1723 ACTTTTCCGCTGGCGCAAGGATCAAGTGTGAGGTGAGCATGAATTTTCAAGAA 1782  
 QY 1797 AGGTGAGCAACAAAGACCCGCGCGGATGACGCGGATTAAGCGAGCCCAAGCGGG 1856

Db 1783 AGGTGAGACCAAGAAAGACCCGCGCGAGTGACGCAATATAGTGAAGCCCAAGCGG 1842  
 QY 1857 CCTGCCCTCAGTCCGCGATCCATCGACGTGACAGCGCGGAAGAGCTCCGCTGACTTG 1916  
 |||||  
 Db 1843 TGGCGAGTCAAGTGGCGGACCATGCAAGTCAAGCGCGGA--AGCTTCATCACTACG 1899  
 QY 1917 CCGACAGGTACCAAAACAAATGTTCTGCTCAGCGCGGCACTTCAAGTGTGTTCCCT 1976  
 |||||  
 Db 1900 CAGACAGGTACCAAAACAAATGTTCTGCTCAGCTGGGCAATGATGCTGTTCCT 1959  
 QY 1977 GCAAGACATGCGAGAGATGAATCAGAAATTTCAACATTTGCTTCAAGCAGCGGAGAGAG 2036  
 |||||  
 Db 1960 GCAAGCAATGCGAGAGATGAATCAGAAATTTCAATATCTGCTCACTCAGGACAGAAAG 2019  
 QY 2037 ACTGTTCAAGTGTCTTCCGCGCTGTGAGAAATCTCAACCGGT--CGTCAAGAAAGAG 2093  
 |||||  
 Db 2020 ACTGTTAAGTGTCTTCC--CGTGTCAAGATCTCAACCGGTTCTGTCTCAAAAGG 2076  
 QY 2094 CGTATCGAAACTCTGTGCAATTCATCATCTGCTGGGCGGCTCCGAGATGCTTCT 2153  
 |||||  
 Db 2077 CGTATCAGAAACTGTGCTACATTCATCATATGGAAGAGTGCCAGA--CGTTCGA 2133  
 QY 2154 CGGCTGCGATCTGTGTCACACGTGGAACCTGATGACTGTGTTCTGAGCAATAATGACTT 2213  
 |||||  
 Db 2134 CTGCTGCGATCTGTCAATGTGATTTGGATGACTGCACTTTGAACAATAATGATTT 2193  
 QY 2214 AAACAGGTATGCTGCGCATGCTTATCTTCCAGATTTGCTGAGAGCAACACTCTGAG 2273  
 |||||  
 Db 2194 AAATCAGTATGCTGCGCATGCTTATCTTCAAGATTTGCTGAGAGCACTCTCTGAA 2253  
 QY 2274 GGCATTCGAGTGTGGGACTTGAACCTGAGGCCCGGAAGCCCAAGCCACAGCAA 2333  
 |||||  
 Db 2254 GGAATPAGACAGTGTGGAAGCTCAAACTGCGCCACCAACCAACCAAGCCGAGAGCGG 2313  
 QY 2334 AAGCAGACAGCGCGGCTGTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2393  
 |||||  
 Db 2314 CATAGGACAGACAGAGGGCTGTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2373  
 QY 2394 GGAATGCAAGGGGAGCGCGCTCAACGCGGAGACGAGCGGCTCGAGCAG--ACAA 2452  
 |||||  
 Db 2374 GGAATGCAAGGGGAGCGCGCTCAACGAGGAGACGCGGCGGCTCGAGCAGTACAA 2433  
 QY 2453 GGCCTACGACCGAGCTCAAGCGGCTGACATCCGTACCTGCGGTATAACCAAGCGCGA 2512  
 |||||  
 Db 2434 AGCCTACGACCGGAGCTCGACAGCGGAGACCAACCGGTACCTCAAGTACCAACCAAGCGCGA 2493  
 QY 2513 CGCCGAGTTTCAAGAGCGCTGTGCAAGAGATACGTCTTTGGGGGCAACCTCGGCGAGC 2572  
 |||||  
 Db 2494 CGCGGAGTTTCAAGAGCGCTTAAAGAGATACGTCTTTGGGGGCAACCTCGGAGCAGC 2553  
 QY 2573 AGTCTTCAAGGCAAGAGCGGCTTCTGAACTCTCGCTGTGTTGAGGAAGCGCTTAA 2632  
 |||||  
 Db 2554 AGTCTTCAAGGCAAGAGCGGCTTCTGAACTCTCGGCTGTGAGGAAGCGCTTAA 2613  
 QY 2633 GACGCTCTTGAAGAAAGCTGCGGTAGAGCAGTGCAGCAAGAGCCAGACTCTCTC 2692  
 |||||  
 Db 2614 GACGCTCTTGAAGAAAGAGCGGCTAGAGCACTCTCTGTGAGCCAGACTCTCTC 2673  
 QY 2693 GGGCATGCGCAAGACAGCGGAGCGGCTTAAAGAGAGTCAATTTGTCAGACTGG 2752  
 |||||  
 Db 2674 GGGATCGGAAAGCGGCGGAGCGCTGCAAGAAAGATTTGTCAGACTGG 2733  
 QY 2753 GCACTCAGATCAGTCCCGGATCCACAACCTCTCGAGAACTTCCAGCAACCCCGCTGC 2812  
 |||||  
 Db 2734 AGACGAGACTCAGTACCTGACCCCAAGCTCTCGGACAGCCACAGAGCCCGCTTGG 2793  
 QY 2813 TGTGGACTTACTAATGCTTCAAGCGGCTGCGGCAACCAATGCGAGCAATTAAGAAAG 2872  
 |||||  
 Db 2794 TCTGGAATAATACATGCTTACAGGCAAGTGGCGCAACCAATGCGAGCAATTAAGAAAG 2853  
 QY 2873 CGCCGAGAGTGGTAAATGCTTCAAGAAATTTGCAATGCGATTCATGCTGGCGGA 2932  
 |||||

Db 2854 CGCGACGGAGTGGGTAATTCCTCCGAAATTGGCAATTGCGATTCCACATGATGGCGCA 2913  
QY 2933 CAGAGTCATCAACCAACGACCCGACCTGGGCTTGGCCACCTACATAACCACTCTA 2992  
Db 2914 CAGAGTCATCAACCAACGACCCGACCTGGGCTTGGCCACCTACATAACCACTCTA 2973  
QY 2993 CAAGCAAAATCTCAGTCTTCAACGGGGGCGCAACGACAACTACTTGGCTACAG 3052  
Db 2974 CAACAAATTTCCAGCCAATCA--GGAGCTTCGAACGACAACTACTTGGCTACAG 3030  
QY 3053 CACCCCTGGGGGTAATTTGATTTCAACAGATTCCACTGCGCACTTTCAACAGTACTG 3112  
Db 3031 CACCCCTGGGGGTAATTTGATTTCAACAGATTCCACTGCGCACTTTCAACAGTACTG 3090  
QY 3113 GCAGGAGTCATCAACAACAATTGGGGATTCGGGCCAAGAGACTCAACTCAAACTCTT 3172  
Db 3091 GCAAGAGTCATCAACAACAATTGGGGATTCGGGCCAAGAGACTCAACTCAAACTCTT 3150  
QY 3173 CAACATCCAAGTCAAGAGGTCACGACGAAATGATGGCGTCAACACCATGCTAATACT 3232  
Db 3151 TAACATTCAGTCAAAAGAGTCAAGAGTCAAGATGACGATACGACGATGCGCAATACT 3210  
QY 3233 TACCAGACGGTCAAGTCTTCTCGACTCGAGTACAGCTTCCGTACGCTCGCTG 3292  
Db 3211 TACCAGACGGTCAAGTCTTCTCGACTCGAGTACAGCTTCCGTACGCTCGCTG 3270  
QY 3293 TGCGACACGAGGCTGCTCCCTCCGTTCCCGGCGGAGCTGTTGATGATCCGCAATACG 3352  
Db 3271 GCGGATCAAGAGTGCCTCCCGCGTTCCAGACAGAGCTTCACTGTCACAGTATGG 3330  
QY 3353 CTACCTGACGCTCAACAATGGCAAGCCGAGCGGTGGAGCTTCACTCTTTTACTGCTG 3412  
Db 3331 ATACCTGACGCTCAACAATGGCAAGCCGAGCTGAGAGTACGCTCTTCACTTTACTGCTG 3390  
QY 3413 ATATTTCCCTTCTCAGATGCTGAGAAACGGGCAACAATTACTTCACTACACTTTGA 3472  
Db 3391 GTACTTCTTCTCAGATGCTGAGAAACGGGCAACAATTACTTCACTACACTTTGA 3450  
QY 3473 GGAAGTCTTCCACAGCAGCTACGCGACAGCCGAGCTGAGCCGCTGATGATCC 3532  
Db 3451 GGAGTCTTCCACAGCAGCTACGCGACAGCCGAGCTGAGCCGCTGATGATCC 3510  
QY 3533 TCTCATGCAACAATACCTGTATTAACCTGAACAAGACTCAAAATCAAGTCCGGAAGTCC 3592  
Db 3511 TCTCATGCAACAATACCTGTATTAACCTGAACAAGACTCAAAATCAAGTCCGGAAGTCC 3570  
QY 3593 AAACAAGACTGTGCTGTTAGCCGCTGCTCCAGCTGGCATGTCTGTCAGCCCAAAA 3652  
Db 3571 GCAGTCAAGGCTTCAAGTTTCTCAGGCGGAGCGAGTGAATGCGGACCAAGTCTAGAA 3630  
QY 3653 CTGCTACCTGGAACCTGTATTCGGCAGACGCGCTTCTTAAACAACAAACAGACAACA 3712  
Db 3631 CTGCTTCTGGAACCTGTATTCGGCAGACGCGCTTCTTAAACAACAAACAGACAACA 3690  
QY 3713 CAACAGCAATTTTACCTGAGCTGGTCTTCAAAATATAATTAACCTCAATGGCGTGAATCC 3772  
Db 3691 CAACAGTGAATTAATCTGAGCTGAGTCAACAAGTACCACTCAATGGCAGAGACTCTCT 3750  
QY 3773 CATCAACCTTGGCACTGTATGGCTTCAACAAGAGCAGAGCAAGTCTTCCAT 3832  
Db 3751 GGTGAATC-----CGGCATGGCAAGCACAAGAGAGATGAAGAAAGTTTTCCTCA 3804  
QY 3833 GAGCGGTGTCATGATTTTGGAAAGAGAGCGCGGAGCTTCAACACTGCAATGGACA 3892  
Db 3805 GAGCGGTGTCATGATTTTGGAAAGAGAGCTCAAGAGAAACAAATGTGAACATTTGAAA 3864  
QY 3893 TGTCAATTAACAGAGAGAGAAATTAAGCCACTAACCTGTGGCCACCGAAAGATT 3952  
Db 3865 GGTCAATTAACAGAGAGAGAAATCGGAAACAACCAATCCCTGGCTACGAGCAGTA 3924  
QY 3953 TGGGACCGTGGCAGTCAATTTCCAGAGCAGACAGACAGACCTGCGACCGGAGATGTGA 4012  
Db 3925 TGGTCTGTATCTACCAACCTCCAGAGAGGCAACAGACAGCAAGCACTACCGCAGATGTCA 3984

QY 4013 TGCTATGGAGACATTACCTGGCATGTGTGGCAAGATAGACAGTGTACCTGCAAGGTCC 4072  
Db 3985 CACACAGGCGTTCTTCCAGGATGGTCTGGCAGACAGAGATGTGTAACCTTCAAGGGCC 4044  
QY 4073 CATTGGGCGCAAAATTCCTCACAAGATGAGACACTTACCCGCTCTCTTATGGCGG 4132  
Db 4045 CATCTGGCAAAAGATTCACACAGGAGAGACATTTTACCCCTCTCCCTCATGGGTG 4104  
QY 4133 CTTTGACTCAAGAACCCGCTCTCAGATCTCATCAAAAACAGCCTGTCTCTCGAA 4192  
Db 4105 ATTGCACTTAACACACCTCTCCACAGATTCTCATCAAGAACACCCCGTACCTGCGAA 4164  
QY 4193 TCCCTCGGCGGAGTTTTCAGCTACAAAGTTTGTCTTCAATTCATCCCAATCTCCACAG 4252  
Db 4165 TCCCTCGACCACTTCAAGTGGCGGAAAGTTTGTCTTCTCATCAACAGTACTCCAGGG 4224  
QY 4253 ACA-AGTGAAGTGAATTTGAATGGGAGCTGCAGAAAGAAACAGCAAGCGCTGGAATC 4311  
Db 4225 ACAGGTCAGCGTGAAGATGAGTGGAGCTGCAGAGAGAAACAGCAACGCTGGAATC 4284  
QY 4312 CCGAAGTCACTACACATCCAAATTAAGCAAAATCTGCCAAGTGAATTTACTGTGACA 4371  
Db 4285 CCGAATTCAGTACACTTCCACTCAACAAGTCTGTTAATGCTGAGCTTACCGTGAAT 4344  
QY 4372 ACAATGAATTTATCTAGAGCTGCGCCCATTTGGACACCCGTTACCTTACCCGCTGT 4431  
Db 4345 CTAAATGCGTGTATTCAGAGCTGCGCCCATTTGGACACGATTAAGTCTGTAATCTGT 4404  
QY 4432 AATTAGTGTAAATCAATAAACCGGTTGATTCGTTTCACTGAACTTGTCTCCGTCC 4491  
Db 4405 AATTGCTGTAAATCAATAAACCGTTAATTCGTTTCACTGAACTTGTCTCCGTCA 4464  
QY 4492 TTCTTATCTATCGGTACAGTGTATAGCTTACACATTAATGCTTGTGCTGCTG 4551  
Db 4465 TTCTTCTTATCTAGTTTCCATGCTACGATGATTAAGTATGATGGCGGTTAATCAAT 4524  
QY 4552 CGATAAAGACTTACGTATCGGGTTACCCCTAGTATGATGAGTTGCCACTCCCTCTG 4611  
Db 4525 ACTACAGGA-----ACCCCTAGTATGATGAGTTGCCACTCCCTCTG 4568  
QY 4612 CGGCTGCTGCTCGGTGGGCGCTGCGGACCAAGGTCGCGACAGCGCAGAGCTCTGCT 4671  
Db 4569 CGGCTGCTGCTCGGTGAGGCGGCGGACCAAGGTCGCGGAGCGGCTTGGCC 4628  
QY 4672 CTGCGGCGCCACCGAGCGAGCGGCGGAGAGAGGAGTGGGCAA 4718  
Db 4629 CGGCGGCGCTCAGTGAAGCGAGCGGCGGAGAGAGGAGTGGCCAA 4675

RESULT 10  
ABA02989  
ID ABA02989 standard; DNA; 4675 BP.  
XX  
AC ABA02989;  
XX  
DT 19-FEB-2002 (first entry)  
XX  
DE Adeno-associated virus 2 genome DNA sequence SEQ ID NO 1.  
XX  
KW Cytostatic; virucide; apoptosis inducer; p53; Saos-2 cell; infection;  
KW cancer; virus; HPV16; HPV18; adeno-associated virus 2; AAV-2; genome; ss.  
XX  
OS Adeno associated virus 2.  
XX  
PN WO200180840-A2.  
XX  
PD 01-NOV-2001.  
XX  
PF 20-APR-2001; 2001WO-GB01795.  
XX  
PR 20-APR-2000; 2000GB-0009887.  
XX



PA (BTGI-) BTG INT LTD.  
XX  
PI Raj K, Beard PM;  
XX  
DR WPI, 2002-041365/05.  
XX

PT Single stranded and/or looped DNA for treating mutant p53 associated  
PT cancer or infection that inhibit cellular p53, having portion with an  
PT base, internally located with respect to any 3' and 5' ends of the DNA  
PT

XX  
PS Disclosure; Page 36-38; 51pp; English.  
XX

CC The invention relates to single stranded and/or looped DNA having a  
CC portion with at least one base, internally located with respect to any 3'  
CC and 5' ends of the DNA, that is unbasepaired with another base in a form  
CC that is capable of being internalised within a target cell, for use in  
CC therapy, with cytostatic and virucide activity. The DNA acts as an  
CC apoptosis inducer in cells that lack p53 functionality, useful for  
CC killing a cell, preferably a dividing cell where the cell is other than a  
CC Sacs-2 cell and the DNA is not configured to express the peptide or  
CC protein that selectively kills the cell. The DNA is preferably in the  
CC form of an AAV or associated with AAV protein which has been treated such  
CC that the DNA is no longer capable of replication or expression in cells  
CC and is associated with or contained within a vehicle which is associated  
CC with one or more viral fibers which facilitate internalisation of the DNA  
CC into a target cell. The DNA is also useful for manufacturing a medicament  
CC for treating an individual suffering from a mutant p53 associated cancer  
CC or an infection that inhibits cellular p53. The method targets cancer  
CC cells or cells infected with p53 inhibiting viruses, such as HPV16 or  
CC HPV18 and only cells that lack p53 activity are killed and no damage to  
CC cellular DNA is involved. The present sequence is that of the  
CC adeno-associated virus (AAV) 2 genomic DNA sequence.  
XX

SQ Sequence 4675 BP, 1198 A; 1262 C; 1251 G; 964 T; 0 other;

Query Match 64.8%; Score 3055.8; DB 24; Length 4675;  
Best Local Similarity 79.9%; Pred. No. 0;  
Matches 3779; Conservative 0; Mismatches 887; Indels 61; Gaps 13;

QY 1 TTGCCACTCCCTCTCTGCGCGCTCGCTCGCTGGGGCGCTCGGACCAAGTCCGC 60  
DB 1 TTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTGAGCGCGGCGACCAAGTCCGC 60  
QY 61 AGACGCGAGAGCTCTGCTCTGCGCGCGCGCGCGCGAGCGAGCGAGAGGAGTG 120  
DB 61 CGACGCGCGCGCTTGGCGCGCGCGCGCGCTCAGTGAAGCGAGCGAGAGGAGTG 120  
QY 121 GGCACTCCATCACTAGGGGTATCGCGAGCGCGCTCCACGCTGCGCGCTCAGCGTGA 180  
DB 121 GCCAACTCCATCACTAGGGGT-----TCCGTGAGGGGTGAGTCTGTGA 163  
QY 181 CGTAATTACGTATAGGG---GAGTGTCTCTGTAATTAAGCTGTACGTAAGTCTTTGC 237  
DB 164 CGTAATTACGTATAGGGGTAGGGAGGTCTGTATTAAGAGTACGTAAGTG-TTTTGC 222  
QY 238 GACATTTTGCAGACCACTGCGCATTTAGGTATATATGCGGAGTGAAGCAGAGAT 297  
DB 223 GACATTTTGCAGACCACTGCTGCGCATTTAGGTATTAAGCCGAGTGAAGCAGAGGT 282  
QY 298 CTCCATTTTG-ACCGCGAAATTGAACGAGCAGACCATGCGGGCTTCTAGAGATCG 356  
DB 283 CTCCATTTTGAAGCGGAGGTTTGAACGCGAGCGCGCATGCGGGGTTTACGAGATTG 342  
QY 357 TGATCAAGGTGCGAGCGAGCTGAGCAGACCTGCGGGCATTTCTGACTCGTTTGTGA 416  
DB 343 TGATTAAGGTCCCGACGAGCCTTGACGCGCATCTGCCGCAATTTCTGACAGCTTTGTGA 402  
QY 417 GCTGGGTGCGCGAGAGGAATGGAGCTGCGCGCGAGTTCTGACATGATCTGAATCTGA 476  
DB 403 ACTGGGTGCGCGAGAGGAATGGAGTGGCGCGCGAGATTCTGACATGATCTGAATCTGA 462  
QY 477 TTGAGCAGGCGACCCCTGACCGGTGCGCGAGAGCTGACGCGCACTTCTGCTCCAATGGC 536

DB 463 TTGAGCAGGCGACCCCTGACCGGTGCGCGAGAGCTGACGCGCACTTCTGACGGAATGGC 522  
QY 537 GCCCGGTGAGTAAGGCCCGGAGGCCCTCTTCTTTGTTCAAGTTCCAGAAGGGCGAGTCT 596  
DB 523 GCCGTGTAGTAAGGCCCGGAGGCCCTTTCTTTGTGCAATTTGAGAAGGAGAGAGT 582  
QY 597 ACTTCACTCCATATTCTGTTGAGAGACCAAGGGGTCAAAATCCATGTGCTGGCGCT 656  
DB 583 ACTTCACTGACAGCTGCTGTGAGAAACACCGGGGTGAATTCATGTTTGGAGCTT 642  
QY 657 TCCGTGAGTCAAGTTAGGAGCAAGCTGTGAGAGCAATCTACCGCGGATCGAGCCGACCC 716  
DB 643 TCCGTGAGTCAAGTTAGGAGCAAGCTGTGAGAGCAATCTACCGCGGATCGAGCCGACTT 702  
QY 717 TGCCCACTGCTTTCGCGGTGAGCAAGCGGTATGCGCGGAGGGGAAACAAGTGG 776  
DB 703 TGCCCACTGCTTTCGCGGTGAGCAAGCGGTATGCGCGGAGGGGAAACAAGTGG 762  
QY 777 TGAGCAGGTGCTACATCCCACTAATCTCTGCGCAAGCTCAGCCGAGCTGAGTGG 836  
DB 763 TGAGTGTGCTACATCCCACTAATCTCTGCGCAAGCTCAGCCGAGCTGAGTGG 822  
QY 837 CGTGAATCAATGAGAGGATATTAAGCGCTGTTGAAGTCCGCGGAGCGCAACGCG 896  
DB 823 CGTGAATCAATGAGAGGATATTAAGCGCTGTTGAAGTCCGCGGAGCGCAACGCGT 882  
QY 897 TCGTGGCGAGCAGCCTGAGCGCGCTGAGCGAGCGAGCGAGAGCAAGAGATCTGA 956  
DB 883 TGTGGCGAGCAGCCTGAGCGCGCTGAGCGAGCGAGCGAGAGCAAGAGATCTGA 942  
QY 957 ACCCAATCTGAGCGCGCTGATCCGCTCAAAACCTCCGCGCTACATGAGCTGG 1016  
DB 943 ATCCCAATCTGATGCGCGCTGATCCGCTCAAAACCTCCGCGCTACATGAGCTGG 1002  
QY 1017 TCGGTGCTGTGAGCGCGCTGATCCGCTCAAAACCTCCGCGCTACATGAGCTGG 1076  
DB 1003 TCGGTGCTGTGAGCGCGCTGATCCGCTCAAAACCTCCGCGCTACATGAGCTGG 1062  
QY 1077 CCTGTACATCTCCTTCAACGCGCTTCCAACTGCGCGTCCAGATCAAGCGCTGTGG 1136  
DB 1063 CCTGTACATCTCCTTCAACGCGCTTCCAACTGCGCGTCCAGATCAAGCGCTGTGG 1122  
QY 1137 ACAATGCGGCAAGATCATGCGCTGACCAATCCGCGCGCTACCTGTAGGCCCG 1196  
DB 1123 ACAATGCGGCAAGATCATGCGCTGACCAATCCGCGCGCTACCTGTAGGCCCG 1182  
QY 1197 CTCCGCGCGGAGCATTAACCAACCGCATCTACCGCATCTGGAAGTGAAGCGCTAGC 1256  
DB 1183 AGCCGCGGAGCATTAACCAACCGCATCTACCGCATCTGGAAGTGAAGCGCTAGC 1242  
QY 1257 AACCTGCTACGCGCGCTCCTTCTCTGCGCTGGCGCCAGAAAGGTTGCGGAAGCGCA 1316  
DB 1243 ATCCCAATATGCGGCTTCCGCTTCTGCGATGGCGCCAGAAAGTTCGCAAGAGGA 1302  
QY 1317 ACACCACTGCTGTTTGGCGCGGCAACCGGCAAGCAACATCGCGGAAGCCATCG 1376  
DB 1303 ACACCACTGCTGTTTGGCGCGGCAACCGGCAAGCAACATCGCGGAAGCCATCG 1362  
QY 1377 CCCAGCGGTGCGCTTCTACGCGCTGCTCACTGAGCAACATGAACTTCCCTCAATG 1436  
DB 1363 CCCAGCGGTGCGCTTCTACGCGCTGCTCACTGAGCAACATGAACTTCCCTCAATG 1422  
QY 1437 ATTGCTGCAACAAGATGTGATCTGTGGAAGAGGCGCAAGATGACCGCCAAAGTCTGG 1496  
DB 1423 ACTGTGCAACAAGATGTGATCTGTGGAAGAGGCGCAAGATGACCGCCAAAGTCTGG 1482  
QY 1497 AGTCCGCAAGGCGCATTTCTGCGCGGAGCAAGGTGCGCTGACCAAAAGTCAAGTCTG 1556  
DB 1483 AGTCCGCAAGGCGCATTTCTGCGCGGAGCAAGGTGCGCTGACCAAAAGTCAAGTCTG 1542  
QY 1557 CCGCCGATGAGCCCGCGCTGATGCTCACTCCAAACCAACATGTGCGCGCTGA 1616

Db 1543 CGGCCAGATAGACCCTGCTCCGCTGATCGTCACTCCAAACCAACATGTGCGCGTGA 1602  
QY 1617 TTGACGGGAACAGACACCACCTTTCGACACAGACGCGCTTGCAGAGCCGGATGTTCAAT 1676  
Db 1603 TTGACGGGAACCTCAACGACCTTTCGACACAGACGCGCTTGCAGAGCCGGATGTTCAAT 1662  
QY 1677 TTGAACCTCAACCGCGCTTTCGACATGACTTTGGCAAGGTGACAAAGCAGAGATCAAG 1736  
Db 1663 TTGAACCTCAACCGCGCTTTCGACATGACTTTGGCAAGGTGACCAAGCAGAGATCAAG 1722  
QY 1737 AGTCTTCCGCTGGCGCAGGATCACTGACCGAGGTGGCGCATGAGTTCTACGTCAAA 1796  
Db 1723 ACTTTTCCGCTGGCGCAGGATCACTGAGTTGAGGTGGAGCATGAATTTCTACGTCAAA 1782  
QY 1797 AGGCTGAGCCCAACAAAGACCCGCGCTTCGATGACCGGATTAAGCCGAGCCCAAGCGGG 1856  
Db 1783 AGGCTGAGCCCAACAAAGACCCGCGCTTCGATGACCGGATTAAGGTGAGCCCAAGCGGG 1842  
QY 1857 CCGCCCTCAGTCCGCGATCGATCGACGTCAGACGCGGAGAGCTCCGGTGAATTTG 1916  
Db 1843 TGGCGAGTCACTTGGCGAGCCATCGACGTCAGACGCGGA--AGCTTCGATCAACTACG 1899  
QY 1917 CGACAGGTACCAAAACAAATGTTCTCGTACGCGGCGATGCTTCAGATGCTTTCCCT 1976  
Db 1900 CAGACAGGTACCAAAACAAATGTTCTCGTACGTCAGTCGATGATGCTTTCCCT 1959  
QY 1977 GCAAGACATGCGAGAGATGAATCAGATTTCAACATTTGCTTCAACGACGAGCAGAG 2036  
Db 1960 GCAGACATGCGAGAGATGAATCAGATTTCAAAATCTGCTTCACTCAGGACAGAAAG 2019  
QY 2037 ACTGTTCAAGTGTCTTCCCGCGCTGTCAGAACTCAACCGGT--CGTCAGAAAGAGA 2093  
Db 2020 ACTGTTAGAGTGTCTTCC--CGTTCAGAACTCAACCGGTCTGCTGTCAGAAAGG 2076  
QY 2094 CGTATCGAAAACCTCTGTCATTCATCTGCTGCGCGGCTCCGAGATGCTTGTCT 2153  
Db 2077 CGTATCGAAAACCTGTCATTCATTCATTCATGGAAGGTGCGAGA--CGCTTGA 2133  
QY 2154 CGGCTGCGATCTGTCACAGTGAAGCTGATGACTGTGTTCTGAGCAATAATGACTT 2213  
Db 2134 CTGCTGCGATCTGTCATTCATTCATTCATGATGACTGATCTTGAACAATAATGATTT 2193  
QY 2214 AAACAGATATGCTGCGATGCTTATCTTCCAGATTTGGCTCGAGACAACTCTGAG 2273  
Db 2194 AAATCAGTATGCTGCGATGCTTATCTTCCAGATTTGGCTCGAGACAACTCTCTGA 2253  
QY 2274 GGCATTCGCGAGTGTGGACTTGAACCTGAGCGCCGAGCCCAAGCCCAACGAGCAA 2333  
Db 2254 GGAATTAAGACAGTGTGGAAGCTCAACCTGCGCCACCAACCAAGCCGAGAGCGG 2313  
QY 2334 AAGCAGACGACGCGCGGCTGCTGCTTCTGCTCAAGTACCTCGGACCTTCAAC 2393  
Db 2314 CATTAAGACGACGAGGGGTCTGCTTCTGCTCAAGTACCTCGGACCTTCAAC 2373  
QY 2394 GGAATCGACAAAGGGGAGCCGCTCAACGCGCGAGCAGCGCGCTCGAGACG-ACAA 2452  
Db 2374 GGAATCGACAAAGGGGAGCCGCTCAACGAGCAGACGCGCGCTCGAGACGTACA 2433  
QY 2453 GGCCTACGACGACGCTCAAGCGGGTGAACATCCGTACCTGGGTATTAACACGCGCA 2512  
Db 2434 AGCTTACGACCGGACGCTCGACGCGGAGACAAACCGTACCTCAAGTACCAACGCGCA 2493  
QY 2513 CGCGGAGTTTCAGGAGCGCTTGAAGAGAGTCAATCCGTACCTGGGTATTAACACGCGCA 2572  
Db 2494 CGCGGAGTTTCAGGAGCGCTTGAAGAGAGTCAATCCGTACCTGGGTATTAACACGAGC 2553  
QY 2573 AGTCTTCCAGGCGCAAGAGCGGCTTTCGAACCTCTCGGTCTGCTGAGGAGGCGCTAA 2632  
Db 2554 AGTCTTCCAGGCGCAAGAGCGGCTTTCGAACCTCTCGGTCTGAGGAGGCGCTTAA 2613  
QY 2633 GACGCTCTGGAAGAAAGCTCCGCTAGACAGTCCGCAACAGACGAGACTCTCTCTC 2692  
Db 2614 GACGCTCTGGAAGAAAGAGCGCGGTAGACACTCTCTCTGAGACGAGACTCTCTCTC 2673

QY 2693 GGCATCGGCAAGACAGCGCCAGCAGCCCGCTAAAGAGACTCAATTTTGGTCAAGCTGG 2752  
Db 2674 GGGAAACCGGAAAGCGCGCCAGCAGCCTGCAAGAAAGATTTGAATTTGGTCAAGCTGG 2733  
QY 2753 CGACTCAGAGTACGTCCTCCGATCCCAACCTCTCGGAGAACCTCCAGCAACCCCGCTGC 2812  
Db 2734 AGACGAGACTCAGTACTGACCCCGAGCCTCTCGGACAGCCACGAGAGCCCTCTGCG 2793  
QY 2813 TGTGGACCTACTACAAATGCTTCAAGCGGTGGCGCAACATGGCAGACAAATTAAGAGG 2872  
Db 2794 TGTGGAACTAATACGATGGCTTACAGCAGTGGCGCAACATGGCAGACAAATTAAGAGG 2853  
QY 2873 CGCCGACGAGTGGGTAAATGCTTCAAGAAATTTGCAATTTGCATTTCCATGCTGGCGCA 2932  
Db 2854 CGCCGACGAGTGGGTAAATTTCTCCGGAATTTGCAATTTGCATTTCCATGATGGCGCA 2913  
QY 2933 CAGAGTCATCACCAACGACCCGACCTGGCGCTTGCCTACATTAACCACTCTA 2992  
Db 2914 CAGAGTCATCACCAACGACCCGACCTGGCGCTTGCCTACATTAACCACTCTA 2973  
QY 2993 CAAGCAATCTCCAGTGTTCACAGGGGGCCAGCAACCAACCTACTTGGCTACAG 3052  
Db 2974 CAACCAATTTCCAGCCAAATCA--GGAGCTCGAAGCAATCACTACTTTGGCTACAG 3030  
QY 3053 CACCCCTGGGGTATTTGATTTCAACAGATTTCCACTTTCACCACTGACTG 3112  
Db 3031 CACCCCTGGGGTATTTGATTTCAACAGATTTCCACTTTCACCACTGACTG 3090  
QY 3113 GCAGGACTCATCAACAAATTTGGGATTTCCGCGCCCAAGAGACTCAACTTCAACTCTT 3172  
Db 3091 GCAAGACTCATCAACAAATTTGGGATTTCCGCGCCCAAGAGACTCAACTTCAACTCTT 3150  
QY 3173 CAACATCCAAATCAAGAGGTCAACGAGATGAGGTCAACCACTCGCTAATTAACCT 3232  
Db 3151 TAACATTCAGTCAAGAGGTCAACGAGATGAGGTCAACCACTCGCTAATTAACCT 3210  
QY 3233 TACCAACAGGTTCAAGTCTTCTCGGACTCGGATACCACTTCCGTCTCGCTC 3292  
Db 3211 TACCAACAGGTTCAAGTCTTCTCGGACTCGGATACCACTTCCGTCTCGCTC 3270  
QY 3293 TGGCACCAGGCTGCTCCCTCCGTTCCCGCGGAGCTGTCATGATTTCCGCAATACGG 3352  
Db 3271 GGGCATCAAGAGTGTCTCCCGCTTCCAGCAGACGCTTTCATGCTGCGCAAGTATGG 3330  
QY 3353 CTACCTGACGCTCAACAAATGGCAGCCAGCCGCTGGAAGCTTCTTACTGCTGGA 3412  
Db 3331 ATACCTCAACCTGAACAAAGGAGTCAAGCAGTGAAGCGCTTCACTTACTGCTGGA 3390  
QY 3413 ATATTTCCCTTCTCAGATGCTGAGAACGGGCAACACTTACTTCACTTACTGGA 3472  
Db 3391 GTAATTTCTTCTCAGATGCTGCTGACCGGAAACAACTTACTTCACTTACTGGA 3450  
QY 3473 GGAAGTCCCTTTCACAGAGCTACGCGCAGAGCCGAGAGCTGAGCGGCTGATGAATCC 3532  
Db 3451 GGAAGTCCCTTTCACAGAGCTACGCGCAGAGCCGAGAGCTGAGCGGCTGATGAATCC 3510  
QY 3533 TCTCATGACCAATACCTGTATTAACCTGAACAGAACTCAAAATCAGTCCGGAAGTCCCA 3592  
Db 3511 TCTCATGACCAATACCTGTATTAACCTGTAGCAGAAACAACTCAAGTGAACCAAC 3570  
QY 3593 AAACAAAGACTTGTGTTAGCCGTGGGTCTCAGCTGGCATGTCTGTTCAAGCCCAAAA 3652  
Db 3571 GCAGTCAAGGCTTCAATTTCTCAGCGGAGCGAGTGAATTTGGAACCACTTGA 3630  
QY 3653 CTGGCTACCTGACCCGTATTCGGCAGCAGCGGCTTCTAAACAAACAAAGACAA 3712  
Db 3631 CTGGCTTCTGACCCGTATTCGGCAGCAGCGGCTTCTAAACAAACAAAGACAA 3690  
QY 3713 CAACAGCAATTTTCACTGACTGTGCTTCAAAATTAACCTCAATGGCGGTGAATCCAT 3772  
Db 3691 CAACAGTGAATTAATCTGAGACTGAGCTACCAAGTACCACTCAATGGCAGAGACTCTCT 3750





QY 138 GGGTAAT-----CGCGAAGCGCCTCCACGCTGCGCGCTC 172  
 Db 124 TGGACTTGCTAGCACTTCTGTATATGCTATACGAAGTTATCCGAGGGGTGG 183  
 QY 173 AGCGCTGACGTAATAATACGTCATAGGG---GAGTGGCTCTGTATTAGCTGTACGCTGAGT 229  
 Db 184 AGTCGTGACGTGAATTAAGTCATAGGTTAGGGAGGTCTGTATTAGAGGTCAAGTGTAGT 243  
 QY 230 GCTTTTGCACTTTTGCGACACCAAGTGGCCATTAGGGTATATATGCGCGAGTGAGCG 289  
 Db 244 G-TTTGCGACATTTTGCGACACCATGTGTCAAGCTGGGTATTAAAGCCGAGTGAGCA 302  
 QY 290 AGCAGATCTCCATTTTG-ACCGGAAATTTGAACGACGACGACGCTGCGGGCTTCTA 348  
 Db 303 CGCAGGGTCTCCATTTTGGAAGCGGGAGGTTGAACGCGCAGCCGCAATGCGGGGTTTA 362  
 QY 349 CGAGATCGTGATCAAGGTGCCGAGCGACCTGGAAGCAGCAGCTGCGGGCATTTCTGACTC 408  
 Db 363 CGAGATTGATTAAGGTCCCGACGACCTTGACGAGCATCTGCCGCAATTTCTGACAG 422  
 QY 409 GTTTGTAGCTGGGTGGCCGAGAAGAAATGGAGCTGCCCGGATTTGACATGATCT 468  
 Db 423 CTTTGTAACTGGGTGGCCGAGAAGAAATGGAGGTGGCCGAGATTTGACATGATCT 482  
 QY 469 GAATCTGATTGACAGGCAACCCCTGACCGGTGGCGAGAAGCTGACGCGGACTTCTGTGT 528  
 Db 483 GAATCTGATTGACAGGCAACCCCTGACCGGTGGCGAGAAGCTGACGCGGACTTCTGTGT 542  
 QY 529 CCAATGGCGCGCGGTGAGTAAGGCGCGGAGCGCCTCTTTGTTGCAATTCGAGAAGGG 588  
 Db 543 GGAATGGCGCGCGGTGAGTAAGGCGCGGAGCGCCTTTTCTTTGTTGCAATTCGAGAAGGG 602  
 QY 589 CGAGTCTTACTTCCACCTCCATATTTCTGTGTGAGAACCAAGGGGTCAAAATTCATGTGTCT 648  
 Db 603 AGAGAGCTTACTTCCACATGACAGCTGTCTGTGAGAAACCAAGGGGTCAAAATTCATGTGTCT 662  
 QY 649 GGGCGCTTCTGAGTCAAGATTAGGGAACAAGCTGTGTGACAGACCATCTACCGGGGATCGA 708  
 Db 663 GGGAGCTTCTGAGTCAAGATTAGGGAACAAGCTGTGTGACAGACCATCTACCGGGGATCGA 722  
 QY 709 GCCGACCTGCGCACTGTGTGCGGTGACCAAGACGCGTAATGCGCGGAGGGGAA 768  
 Db 723 GCCGACCTTTCGCAAACTGTGTGCGGTGACCAAGACGCGTAATGCGCGGAGGGGAA 782  
 QY 769 CAAGGTGTGACGAGTGTACATCCCACTAATCTCTGCGCAAGCTCAGCCGAGCT 828  
 Db 783 CAAGGTGTGACGAGTGTACATCCCACTAATCTCTGCGCAAGCTCAGCCGAGCT 842  
 QY 829 GCAGTGGCGGTGACCTAATGAGGAGTATTAAGCGGCTGTGTAACCTGCGCGAGCG 888  
 Db 843 CCAGTGGCGGTGACCTAATGAGGAGTATTAAGCGGCTGTGTAACCTGCGCGAGCG 902  
 QY 889 CAAACGGCTGTGGCGCAGACCTGACCAAGTCAAGCAGACCAAGAGCAGAAACAAGGA 948  
 Db 903 TAAACGGTGTGGCGCAGACCTGACCAAGTCAAGCAGACCAAGAGCAGAAACAAGGA 962  
 QY 949 GAATCTGAACCCCAATTTGACGCGCTGTCAATCCGCTCAAAAACTTCCGCGCTACAT 1008  
 Db 963 GAATCAGAAATCCCAATTTGATGCGCGGTGATCAGATCAAAAACTTCAAGCAGTACAT 1022  
 QY 1009 GAGGCTGTGGGTGGGTGTGAGACGCGGCGATCACCTCCGAGAAGCAGTGATCCAGGA 1068  
 Db 1023 GAGGCTGTGGGTGGGTGTGAGACGCGGCGATCACCTCCGAGAAGCAGTGATCCAGGA 1082  
 QY 1069 GAGCAGGCGCTGTACATCTCTTCAAGCGCGCTTCCCACTGCGCGTCCAGATCAAGGC 1128  
 Db 1083 GAGCAGGCGCTGTACATCTCTTCAAGCGCGCTTCCCACTGCGCGTCCCAATCAAGGC 1142  
 QY 1129 CGCTCTGACAAATGCGCGCAAGATCATGGCGCTGACCAAAATCCGCGCGGACTACCTGT 1188  
 Db 1143 TGCTTTGACAAATGCGCGCAAGATCATGGCGCTGACCAAAATCCGCGCGGACTACCTGT 1202  
 QY 1189 AGGCGCGCTGCGCGCGGAGCATTAACAACGAGCATCTACCGCATCTGTGAGCTGAA 1248

Db 1203 GGGCCAGCAGCCCGTGAGGAGCATTTCCAGCAATCGGATTTATAAATTTTGAACATAAA 1262  
 QY 1249 CGGCTACGAACCTGCTACGCGCGCTCGCTTCTCTCGGCTGGGCCAGAAAAAGTTCCG 1308  
 Db 1263 CGGCTACGAACCTGCTACGCGCGCTCGCTTCTCTCGGCTGGGCCAGAAAAAGTTCCG 1322  
 QY 1309 GAAGCGCAACCATCTGCTGCTGTTGGGCGCGCCACCAAGGCGCAAGCAATCGCGGA 1368  
 Db 1323 CAAGAGGAACACCATCTGCTGCTGTTGGGCGCTGCAACTACCGGGAAGACCAATCGCGGA 1382  
 QY 1369 AGCCATGCGCCACGCGCGTCTGCGCTGCGTCAACTGACCAATGAGAACTTCC 1428  
 Db 1383 GGCATAGCCACACTGTGCCCTTCAAGGGTGCGTAACCTGACCAATGAGAACTTCC 1442  
 QY 1429 CTTCAATGATTCGCTGACAAAGATGTGATCTGTGTGGAGAGGCGCAAGATGACGGCCAA 1488  
 Db 1443 CTTCAACGACTGTGTGACAAAGATGTGATCTGTGTGGAGAGGCGCAAGATGACGGCCAA 1502  
 QY 1489 GGTGTGAGTTCGCCAAGGCGCATTTCTGCGCGGACGAAAGTGTGACCAAAAGTG 1548  
 Db 1503 GGTGTGAGTTCGCCAAGGCGCATTTCTGAGAGAAAGTGTGACCAAAAGTG 1562  
 QY 1549 CAAGTGTGCGCCGACATGACCCCAACCCCGTGTGATGTCACTTCAACCAACATGTG 1608  
 Db 1563 CAAGTGTGCGCCGACATGACCCCAACCCCGTGTGATGTCACTTCAACCAACATGTG 1622  
 QY 1609 CGCGGTGATGACGGAACACGACCACTTTCGAGCACCAGCAGCGTTCGAGACCGGAT 1668  
 Db 1623 CGCGGTGATGACGGAACACGACCACTTTCGAGCACCAGCAGCGTTCGAGACCGGAT 1682  
 QY 1669 GTTCAATTTGAATCACTACCCCGCGTCTGAGCATGACTTTGGAAGGTGACAAAGCAGGA 1728  
 Db 1683 GTTCAATTTGAATCACTACCCCGCGTCTGAGCATGACTTTGGAAGGTGACAAAGCAGGA 1742  
 QY 1729 AGTCAAGAGTCTTCTCGCTGGGCGCAGATCAAGTGAACCAAGTGGCGCATGATTTCTA 1788  
 Db 1743 AGTCAAGAGTCTTCTCGCTGGGCGCAGATCAAGTGAACCAAGTGGCGCATGATTTCTA 1802  
 QY 1789 CGTCAAGAAAGGCTGAGGCCAACAAGAACCCGCGCGCGATGACGCGATTAAGCGAGCC 1848  
 Db 1803 CGTCAAGAAAGGCTGAGGCCAACAAGAACCCGCGCGCGATGACGCGATTAAGCGAGCC 1862  
 QY 1849 CAAGCGGCGCTGCGCGCTCAGTCCGATTCATGACGTGACAGCGGAGAGAGCTCCGCT 1908  
 Db 1863 CAAGCGGCGCTGCGCGCTCAGTCCGATTCATGACGTGACAGCGGAGAGAGCTCCGCT 1919  
 QY 1909 GAGCTTTGCCGACAGTACCAAAACAATGTCTCTGTCACGCGGCGCATGTTCAGATGCT 1968  
 Db 1920 CAACTACGACAGACAGTACCAAAACAATGTCTCTGTCACGCGGCGCATGTTCAGATGCT 1979  
 QY 1969 GTTTCCTGCAAGACATGCGAGAGATGAATCAGAAATTTCAATTTGCTTCAAGCAGCG 2028  
 Db 1980 GTTTCCTGCAAGACATGCGAGAGATGAATCAGAAATTTCAATTTGCTTCAAGCAGCG 2039  
 QY 2029 GACGAGAGACTGTTCAAGTGTCTTCCCGGCGGTGACAAATCTCAACCGGT--CGTCAAG 2085  
 Db 2040 ACAGAAAGACTGTTAGAGTGTCTTCC--CGTCAAGAAATCTCAACCGGTGTCTGTCTGT 2096  
 QY 2086 AAAGAGAGCTATCGAAATCTGTGCCATTCATCATCTGTGGGCGGCGCTCCCGAGAT 2145  
 Db 2097 CAAAGAGCGGTATCAGAAATCTGTCTACATTCATCATCTGTGGGCGGCGCTCCCGAGAT 2153  
 QY 2146 TGCTTCTGCGCGCTGCGATGTGTCAAGTGAAGTGAAGTGTGTTCTGAGCAATA 2205  
 Db 2154 CGCTTGCACTGCGCTGCGATGTGTCAAGTGAAGTGAAGTGTGTTCTGAGCAATA 2213  
 QY 2206 AATGACTTAACAGATAGTGTGCGGATGTGTTATCTTCCAGATTTGCTGAGGACAAAC 2265  
 Db 2214 AATGATTTAATCAGATAGTGTGCGGATGTGTTATCTTCCAGATTTGCTGAGGACAAAC 2273  
 QY 2266 TCTGTAGGGGATTCGCGAGTGTGGACTTGAACCTGAGCCCGGAGCGGCAAGCCCA 2325

Db 2274 TCTGTGAAGGAATAAGACAGTGTGGAAGCTCAAACTGGCCACACCAACCAAAAGCCCG 2333  
QY 2326 ACCAGCAAAAGCAGAGCAGCGCGGGGTCTGGTCTCTCTGCTACAAAGTACCTCGAGC 2385  
Db 2334 CAGAGCGCATAAAGACGACAGAGGGGTCTGTGCTTCTGTGTACAAAGTACCTCGAGC 2393  
QY 2386 CTTCAACGCACTCGACAAGGGGAGCCCGTCAACGCGCGGACGACGAGCGCCCTCGAGC 2445  
Db 2394 CTTCAACGCACTCGACAAGGGAGAGCCGGTCAACGAGGACAGACGCCGGCCCTCGAGC 2453  
QY 2446 ACACAAGGCTTACGACGACGAGCTCAAGCGGGTGAACAATCCGTACCTGCGGTATAAC 2505  
Db 2454 ACACAAGGCTTACGACCGGCACTCGACAGCGGAGACAACCCGTACCTCAAGTACAACC 2513  
QY 2506 ACGCCGACGCCGAGTTTCAAGAGCGCTCGCAAGAGATAAGTCTTTTGGGGCAACCTCG 2565  
Db 2514 ACGCCGACGCCGAGTTTCAAGAGCGCCCTTAAGAGATAAGTCTTTTGGGGCAACCTCG 2573  
QY 2566 GCGGAGCAGTCTTCCAGGCCAAGAGCGGGTCTCGAACTCTCGGTCTGGTTGAGGAAG 2625  
Db 2574 GACGAGCAGTCTTCCAGGCCAAGAGAGGGTCTTGAACTCTGGGCTGTGTTGAGGAAC 2633  
QY 2626 GCGTAAGACGGCTCTGGAAGAAAGCTCCGGTAGACAGTCGCCACAGAGCCAGACT 2685  
Db 2634 CTGTTAAGACGGCTCCGGGAAAAAAGAGCGCGGTAGACACTCTCTGTGAGCCAGACT 2693  
QY 2686 CCTCCTCGGCATCGGCAAGACAGGCCAGACCCGCTAAAGAGAGACTCAATTTTGTGTC 2745  
Db 2694 CCTCCTCGGGAACCGGAAGCGGGGCCAGCAGCCTGCAAGAAAAAGATTGTAATTTGTGTC 2753  
QY 2746 AGACTGGCGACTCAAGTCAGTCTCCCGATCCACAACCTCTCGAGAGACCTCCAGCAACC 2805  
Db 2754 AGACTGGAGACCGCAAGCTCAGTACCTGACCCCGCTCTCGAGACGCCACAGCAGCC 2813  
QY 2806 CCGTGTGTGGGACTTAACAATGGCTTCAGGCGGTGGCGCAACCAATGGCAGACAATA 2865  
Db 2814 CCTGTGTGGGAACTAATAAGATGGCTACAGGCAGTGGCGCAACCAATGGCAGACAATA 2873  
QY 2866 ACGAAGCGCGCAGCGAGTGGGTATGCTCAGGAAATTGGCATTGGATTCCACATGGC 2925  
Db 2874 ACGAGGCGCGCAGCGAGTGGGTATTCCTCGGAAATTGGCATTGGATTCCACATGGA 2933  
QY 2926 TGGGCGACAGATCATCAACCAAGCAGCAGCCGCACTGGGCTTGGCCCACTCAATAACC 2985  
Db 2934 TGGGCGACAGATCATCAACCAAGCAGCAGCCGCACTGGGCTTGGCCCACTCAATAACC 2993  
QY 2986 ACCTCTACAAGCAAAATCTCCAGTGTCTCAACGGGGGCCAGCAAGCAACAACAATACTCG 3045  
Db 2994 ACCTCTACAAGCAAAATTTCCAGCCAATCA---GGAGCCTCGAAGCAACAATCACTATTG 3050  
QY 3046 GCTACAGCACCCCTGGGGGTATTTGATTTCACAGATTCCACTGCCACTTTTCCACCAC 3105  
Db 3051 GCTACAGCACCCCTTGGGGGTATTTGACTTCACAGATTCCACTGCCACTTTTCCACCAC 3110  
QY 3106 GTGACTGGCAGGCACTCATCAACAACAATTGGGGATTCCGGGCCCAAGAGACTCAACTTCA 3165  
Db 3111 GTGACTGGCAAAAGACTCATCAACAACAATGGGGATTCCGACCCCAAGAGACTCAACTTCA 3170  
QY 3166 AACTCTTCAACATCAAGTCAAGGAGTCAAGATGATGGCTCAACAACATCGCTA 3225  
Db 3171 AGCTCTTAAATTCAGTCAAAAGAGTCAAGCAGAAATGACGATACGACGCAATTGGCCA 3230  
QY 3226 ATAACCTTACGACGAGTTCAAGTCTTCTCGGACTCGGAGTACAGGTTCCGTACGTCC 3285  
Db 3231 ATAACCTTACGACGAGTTCAAGTCTTCTCGGACTCGGAGTACAGGTTCCGTACGTCC 3290  
QY 3286 TCGGCTCTGCGCACGAGGGCTGCTCCCTCCGTTCCCGGCGAGCTGTTCAATGATTCGCG 3345  
Db 3291 TCGGCTCGGCGCATCAAGGATGCTCCCGCGCTTCCCAAGACGCTCTTCAATGATGCTCCAC 3350  
QY 3346 AATACGGCTACCTGAGCTCAACAATGGCAGCCAAAGCCGTGGAGCTTCACTTTTACT 3405  
Db 3351 AGTATGATACCTCAACCTGAACAACGGGAGTCAAGGCAATGAGACGCTCTTCAATTTTACT 3410

QY 3406 GCCTGGAATATTTCCCTTCTCAGATGCTGAGAACGGGCAACAATTACCTTACGCTACA 3465  
Db 3411 GCCTGGAATATTTCCCTTCTCAGATGCTGCGTACCAGAAACAATTACCTTACGCTACA 3470  
QY 3466 CCTTTGAGGAAGTCCCTTTCACAGACGCTACGCGCACAGCCAGAGCCCTGAGCCGGCTGA 3525  
Db 3471 CTTTGAAGAGCTTCTTTCACAGACGCTACGCTCACAGCCAGAGTCTGAGCCGTCTCA 3530  
QY 3526 TGAATCTCTCATGACCAATACCTGTATTAACCTGAACAGAACTCAAAATCAGTCCGAA 3585  
Db 3531 TGAATCTCTCATGACCAAGTACCTGTATTAACCTGAGCAGAACAACTCAAGTGA 3590  
QY 3586 GTGCCCAAAACAAGACTGTGTTTACCCGTGGGTCTCCAGCTGGCATGTCTGTTCAGC 3645  
Db 3591 CCACACGCACTCAAGCTTCAGTTTCTCAGGCCGAGCGAGTGAATTCGGAGCCAGT 3650  
QY 3646 CCAAAATGGCTACCTGACCCCTGTATCGGACAGCGGCTTTTAAACAAAACAG 3705  
Db 3651 CTAGGAATGGCTTCTCGACCCCTGTATCCGACAGCAGGATATCAAGACATCTGCGG 3710  
QY 3706 ACAACAACAAGCAATTTTACCTGGACTGTGCTTCAAAATATAACCTCAATGGGCGTG 3765  
Db 3711 ATAACAACAACAAGTAAATCTCGTGAAGCTGAGCTACCAAGTACCACTCAATGCGAGAG 3770  
QY 3766 AATCCATCATCAACCTTGGCACTGTATGGCTTCAACAACAAGCAGAGCAAGCAAGTCT 3825  
Db 3771 ACTCTGTGTGAATCCGGGCCCGCATGCGAACCCACAAGCAGATGAAGAAAGTTT 3830  
QY 3826 TTCCCATGAGCGGTGTATGATTTTGGAAAAAGAGCGCGGAGCTTCAAAACATGCTCAT 3885  
Db 3831 TTCCCATGAGCGGGTGTATGATTTTGGAAAAAGAGCGCGGAGCTTCAAAACATGCTCAT 3890  
QY 3886 TGAACAATGTCATGATTACAGACGAGAAATTAAGCCATAACCTGTGGCCACCG 3945  
Db 3891 TTGAAGAAGTCATGATTACAGACGAGAAATCAGACCAACCAATCCGTGCTACGG 3950  
QY 3946 AAAAATTTGGGACCGGTGCAATTTCCAGAGCAGACAGACAGACCCCTGCAACCGAG 4005  
Db 3951 AGCAGATGCTTCTGTATCTACCAACCTCCAGAGAGGCAACAGACAGTACCGCAG 4010  
QY 4006 ATGTGATGCTATGGAGCATTAACCTGGCATGTGTGGCAAGATAGAGAGTGTACTGTC 4065  
Db 4011 ATGTGATGCTATGGAGCATTAACCTGGCATGTGTGGCAAGATAGAGAGTGTACTGTC 4070  
QY 4066 AGGTGCCATTTGGGCCAAATTCCTCACACAGATGAGACACTTCAACCCGTCTCTCTTA 4125  
Db 4071 AGGGGCCATCTGGGCAAGATTCACACAGGACGACATTTCAACCCCTCTCTCTCA 4130  
QY 4126 TGGCGGCTTGGACTCAAGAACCCGCTCTCAGATCTCATCAAAAAACAGCGCTGTTTC 4185  
Db 4131 TGGGTGATTCGGAATTAACACCTCTCTCACAGATTCATCAAGAACACCCCGTAC 4190  
QY 4186 CTGCAATCTCCGGCGAGTTTCAAGTACCAAAAGTTTGTCTCATTCACCCCAATACT 4245  
Db 4191 CTGCAATCTCTCGACCACTTCAGTGGCGCAAAAGTTTGTCTCTCATCACACAGTACT 4250  
QY 4246 CCACAGACAAGTGAAGTGTGAATTTGAATGGAGCTGCAAGAAAGAAACAGCAAGCGCT 4305  
Db 4251 CCACAGACAAGTGAAGTGTGAATTTGAATGGAGCTGCAAGAAAGAAACAGCAAGCGCT 4310  
QY 4306 GGAATCCGGAAGTGAAGTGAATTTGAATGGAGCTGCAAGAAATCTGCCAAGTGTGATTTACTG 4365  
Db 4311 GGAATCCGGAATTCAGTACCTTCAACTACCAACAAGTGTGTTAATGTGACTTTACTG 4370  
QY 4366 TGGACAACAATGACTTTATATGAGCCTGCCCCATTTGGCAACCGTTACCTTACCGCTC 4425  
Db 4371 TGGACAACAATGAGCTGTATGAGAGCCTGCCCCATTTGGCAACAGATACCTGACTGTA 4430  
QY 4426 CCCTGTAATTAAGTGTATTAATCAATAAACCGGTGATTCGTTCAAGTGAACCTTGTGCTC 4485  
Db 4431 ATCTGTAATTTGCTTGTATTAATCAATAAACCGTTAATTCGTTCAAGTGAACCTTGTGCTC 4490

OY 4486 CTGTCCTTCTATCTATC 4504  
DB 4491 TGGGTATTTCTTCTATC 4509  
RESULT 12  
AAD00834  
ID AAD00834 standard; DNA; 8178 BP.  
AC AAD00834;  
XX  
DT 08-SEP-2000 (first entry)  
XX  
DE Chimeric adeno-associated virus 2/3 capsid encoding helper plasmid.  
XX  
KW Adeno-associated virus; AAV2; AAV3; chimeric; gene therapy; plasmid;  
KW recombinant parvoviral vector; cellular tropism; rep gene; cap protein;  
KW capsid; gene delivery; VP1; VP2; VP3; ss.  
XX  
OS Chimeric - Adeno associated virus serotype 2.  
OS Chimeric - Adeno associated virus serotype 3.  
XX  
FH Key  
FT CDS  
FT 251..2180  
FT /product= a  
FT /note= "Rep 68"  
FT encodes Rep 68 protein of AAV2. However, the stop  
FT position given in the specification for this CDS does  
FT not correspond to stop codon"  
FT 251..2114  
FT /product= b  
FT /note= "Rep 78"  
FT encodes Rep 78 protein of AAV2. However, the stop  
FT position given in the specification for this CDS does  
FT not correspond to stop codon"  
FT 923..2114  
FT /product= c  
FT /note= "Rep 52"  
FT encodes Rep 52 protein of AAV2. However, the stop  
FT position given in the specification for this CDS does  
FT not correspond to stop codon"  
FT 923..2180  
FT /product= d  
FT /note= "Rep 40"  
FT encodes Rep 40 protein of AAV2. However, the stop  
FT position given in the specification for this CDS does  
FT not correspond to stop codon"  
FT 2133..4342  
FT /product= e  
FT /note= "VP1 cap protein"  
FT encodes VP1 protein of AAV2. However, the stop  
FT position given in the specification for this CDS does  
FT not correspond to stop codon"  
FT 2544..4342  
FT /product= f  
FT /note= "VP2 cap protein"  
FT encodes VP2 protein of AAV2. However, the positions  
FT given in the specification for this CDS do not correspond  
FT to start or stop codons"  
FT 2739..4342  
FT /product= g  
FT /note= "VP3 cap protein"  
FT encodes VP3 protein of AAV2 in which loops 2-4 are  
FT replaced with the corresponding region from AAV3.  
FT However, the stop position given in the specification  
FT for this CDS does not correspond to stop codon"

FT misc\_feature 3184..4092  
FT /product= h  
FT /note= "AAV3 sequence"  
XX  
PN WO200028004-A1.  
XX  
PD 18-MAY-2000.  
XX  
PF 10-NOV-1999; 99WO-US26505.  
XX  
PR 10-NOV-1998; 98US-0107840.  
PR 10-MAR-1999; 99US-0123651.  
XX  
PA (UYN-) UNIV NORTH CAROLINA.  
XX  
PI Rabinowitz JE, Samulski RJ, Xiao W;  
XX  
DR WPI; 2000-376523/32.  
XX

PT Recombinant parvoviral vectors with altered packaging, tropisms and  
PT immunogenic properties, useful in gene therapy protocols -  
XX  
PS Example 27; Page 143-147; 153pp; English.

CC The patent discloses modified parvovirus vectors with advantageous  
CC antigenic properties, packaging capabilities and cellular tropisms.  
CC These vectors can be used in standard recombinant DNA protocols e.g. gene  
CC therapy for delivering nucleic acids to cells. The present sequence is  
CC a helper plasmid encoding a chimeric adeno-associated virus serotype 2/3  
CC (AAV2/3) capsid. The sequence contains the AAV2 rep coding sequences,  
CC most of the AAV2 capsid coding sequences with the exceptions that  
CC loops 2-4 from the AAV2 VP3 subunit is replaced with the corresponding  
CC region from AAV3, in a Bluescript backbone. The rep  
CC coding sequence encodes four proteins responsible for replication,  
CC Rep 68, Rep 78, Rep 52 and Rep 40, and the capsid (cap) coding region  
CC encodes three structural proteins VP1, VP2 and VP3.  
CC Recombinant parvovirus comprising the chimeric capsid is useful  
CC for gene delivery.

XX  
SQ Sequence 8178 BP; 2082 A; 2078 C; 2027 G; 1991 T; 0 other;

Query Match 62.4%; Score 2941.8; DB 21; Length 8178;  
Best Local Similarity 81.2%; Pred. No. 0;  
Matches 3517; Conservative 0; Mismatches 792; Indels 20; Gaps 8;

OY 205 GTCCGTATTAGCTGTACAGTGTGCTTTTGGACATTTGGACACACCGTGGCAATT 264  
DB 121 GTCCGTATTAGAGTCAAGTGTG-TTTTGGACATTTGGACACACCATGTGTACG 179  
OY 265 TAGGTATATATGCGGAGTGAAGCAGCAGATCTCCATTTTG-ACCGGAATTTGAC 323  
DB 180 CTGGTATTTAAGCCCGAGTGAAGCAGCAGGCTCTCCATTTGAAGCGGAGTTGAAC 239  
OY 324 GAGCAGCAGCCATGCGGCTTTCTACGAGATCGTATCAAGGTGCCGAGCCTTGACG 383  
DB 240 GCGCAGCCGCGATGCGGGGTTTACGAGATTGTATTAAGTCCCGCAGCAGCCTTGACG 299  
OY 384 AGCAGCTGCGGCGCATTTCTGACTGTTGTGAGCTGGGTGGCCGAGAAGGAATGGAGC 443  
DB 300 GGCATCTGCGCGCATTTCTGACAGCTTTGTGAAGTGGGTGGCCGAGAAGGAATGGAGT 359  
OY 444 TGCCCCCGGATTTCTGACATGATCTGATTTGACAGGACCCCTGACCGTGGCGG 503  
DB 360 TGCCGCCAGATTCTGACATGATCTGATTTGACAGGACCCCTGACCGTGGCGG 419  
OY 504 AGAAGCTGACGCGGCACTTCTGCTCCATATGCGCGCGGTGAAGAGCCCGAGGCGCC 563  
DB 420 AGAAGCTGACGCGGCACTTCTGACGGAATGCGCGCGGTGAAGAGCCCGAGGCGCC 479  
OY 564 TCTTCTTTTGTTCAGTTCGAGAAGGGCGAGTCTTCCATCTCCATATTTCTGTGAGA 623  
DB 480 TTTTCTTTTGTTCAGTTCGAGAAGGGCGAGTCTTCCATCTCCATATTTCTGTGAAA 539







FT	/product= "Rep 78"	
FT	/note= "The specification states that this region	
FT	encodes Rep 78 protein of AAV2. However, the stop	
FT	position given in the specification for this CDS does	
FT	not correspond to stop codon"	
FT	923..2114	
FT	/*tag= c	
FT	/product= "Rep 52"	
FT	/note= "The specification states that this region	
FT	encodes Rep 52 protein of AAV2. However, the stop	
FT	position given in the specification for this CDS does	
FT	not correspond to stop codon"	
FT	923..2180	
FT	/*tag= d	
FT	/product= "Rep 40"	
FT	/note= "The specification states that this region	
FT	encodes Rep 40 protein of AAV2. However, the stop	
FT	position given in the specification for this CDS does	
FT	not correspond to stop codon"	
FT	2133..4315	
FT	/*tag= e	
FT	/product= "Capsid protein VP1"	
FT	/note= "The specification states that this region	
FT	encodes VP1 protein of AAV2. However, the stop	
FT	position given in the specification for this CDS does	
FT	not correspond to stop codon"	
FT	2544..4315	
FT	/*tag= f	
FT	/product= "VP2 cap protein"	
FT	/note= "The specification states that this region	
FT	encodes VP2 protein of AAV4. However, the positions	
FT	given in the specification for this CDS do not correspond	
FT	to start or stop codons"	
FT	2724..4315	
FT	/*tag= g	
FT	/product= "VP3 cap protein"	
FT	/note= "The specification states that this region	
FT	encodes VP3 protein of AAV2. However, the stop	
FT	position given in the specification for this CDS does	
FT	not correspond to stop codon"	
PN		
XX	WO200028004-A1.	
XX		
XX	18-MAY-2000.	
XX		
PF	10-NOV-1999; 99WO-US26505.	
XX		
XX	10-NOV-1998; 98US-0107840.	
PR	10-MAR-1999; 99US-0123651.	
XX		
PA	(UVNC-) UNIV NORTH CAROLINA.	
XX		
PI	Rabinowitz JE, Samulski RJ, Xiao W;	
XX		
DR	WPI; 2000-376523/32.	
XX		
PT	Recombinant parvoviral vectors with altered packaging, tropisms and	
PT	immunogenic properties, useful in gene therapy protocols -	
XX		
PS	Example 20; Page 135-139; 153pp; English.	
XX		
XX	The patent discloses modified parvovirus vectors with advantageous	
CC	antigenic properties, packaging capabilities and cellular tropisms.	
CC	These vectors can be used in standard recombinant DNA protocols e.g. gene	
CC	therapy for delivering nucleic acids to cells. The present sequence is	
CC	a helper plasmid encoding a chimeric adeno-associated virus serotype 2	
CC	(AAV2) capsid in which the AAV4 VP2 is substituted. The sequence contains	
CC	the AAV2 rep coding sequences, most of the AAV2 VP1 and VP3 coding	
CC	sequences, the entire AAV4 VP2 coding sequence and some of the	
CC	AAV4 VP1 and VP3 coding sequences in a pBluescript backbone. The rep	
CC	coding sequence encodes four proteins responsible for replication,	
CC	Rep 68, Rep 78, Rep 52 and Rep 40 and the VP1, VP2 and VP3 are the	
CC	structural proteins encoded by capsid (cap) coding region.	

CC	Recombinant parvovirus comprising the chimeric capsid is useful for gene delivery.
CC	Sequence 8151 BP; 2048 A; 2095 C; 2047 G; 1961 T; 0 other;
XX	Query Match 57.9%; Score 2730; DB 21; Length 8151;
CC	Best Local Similarity 78.5%; Pred. No. 0;
XX	Matches 3395; Conservative 0; Mismatches 890; Indels 41; Gaps 9;
QY	205 GTCTGTATTAGCTGTACAGTGAAGTCTTTGGACATTTTGGACACCAAGTGGCCATT 264
DB	121 GTCTGTATTAGAGTCAAGTGAAGT-TTTGGACATTTTGGACACCAAGTGGTCAAG 179
QY	265 TAGGGTATATATGCGCGAGTGAGCGAGCAGATCTCCATTGTG-ACCGCAAAATTGAAC 323
DB	180 CTGGTATTAAAGCCGAGTGAGCAGCAGGCTCTCATTTTGAAGCGGAGGTTTGAAC 239
QY	324 GAGCAGCAGCCATGCGCGGCTTCTACAGATCTGATCAAGGTGCCGAGCAGCTTGACG 383
DB	240 GCGCAGCCGCAATGCGGGGTTTACGAGATTGTAAAGTCCCAAGCACTTGACG 299
QY	384 AGCAGCTGCGCGCATTTCTGACTCGTTTGTAGCTGGGTGCCGAGAAAGATGGAGC 443
DB	300 GGCATCTGCCCGCATTTCTGACAGCTTTGTGAAGTGGGTGCCGAGAAAGATGGAGT 359
QY	444 TGCCCCCGGATTTGACATGATCTGAATCTGAATTGAGCAGCAGCACCCTGACCGTGCCG 503
DB	360 TGCCCGCAGATTTGACATGATCTGAATCTGAATTGAGCAGCAGCACCCTGACCGTGCCG 419
QY	504 AGAAGCTGACGCGGCACTTCTGTGTCCAATGCGCGCGCTGAGTAAGGCCCGCGAGGCC 563
DB	420 AGAAGCTGACGCGGCACTTCTGTGTCCAATGCGCGCGCTGAGTAAGGCCCGCGAGGCC 479
QY	564 TCTTCTTTGTTCAAGTTCGAGAAAGGCGAGTCTTCACTTCCACCTCCATATTCTGTGGAGA 623
DB	480 TTTTCTTTGTTCAAGTTCGAGAAAGGCGAGTCTTCCACATGCAAGTCTGTGGAGAA 539
QY	624 CCACGGGGGTCAAAATCCATGTGTGGCGCGCTTCTCTGAGTCAGATTAGGCAAGCTGG 683
DB	540 CCACGGGGGTCAAAATCCATGTGTGGCGCGCTTCTCTGAGTCAGATTGCGGAAAACTGA 599
QY	684 TGCAGACCATCTACCGCGGGATCGAGCCGACCCCTGCCCCAACTGTTGCGGTGACCAAGA 743
DB	600 TTCAGAGATTTCACCGCGGGATCGAGCCGACTTGCACAACTGTTGCGGTGACCAAGA 659
QY	744 CGCGTAATGCGCGCGAGGGGGGAAACAAGTGTGACGAGTGTACATCCCAACTACC 803
DB	660 CCAGAAATGCGCGCGAGGGGGGAAACAAGTGTGATGAGTGTACATCCCAATTAAT 719
QY	804 TCCTGCCCAAGACTCAGCCCGAGCTGCAGTGGGCGTGACTAACATGAGAGATATATAA 863
DB	720 TGCTCCCAAAATCCAGCTGAGCTCCAGTGGGCGTGACTAATATGAGACAGTATTTAA 779
QY	864 GCGCCTGTTGAACCTGCGCGAGCGCAAAAGGCTGTTGGCGCAGCACCTGACCAAGTCA 923
DB	780 GCGCCTGTTGAATCTCAACGAGCGTAAACGGTTGGTGGCGCAGCATCTGACGCACGTGT 839
QY	924 GCCAGACCCAGAGAGCAAGCAAGGAGATCTGAACCCCAATTTGACGCGGCTGTCAATCC 983
DB	840 CGCAGACGAGAGAGCAAGCAAGGAGATCTGAATCCCAATTTGATGCGCGGTGATCA 899
QY	984 GGTCAAAAACCTCGCGCGCTACATGAGCTGTGGGTGGCTGTGTGACCGGGGATCA 1043
DB	900 GATCAAAAACCTCGCGCGCTACATGAGCTGTGGGTGGCTGTGTGACCAAGGGAATTA 959
QY	1044 CCTCCGAGAAGCAGTGGATCCAGAGAGCAAGGCTCTGATCATCTCTTCAAGCGCGCTT 1103
DB	960 CCTCCGAGAAGCAGTGGATCCAGAGAGCAAGGCTCTATACATCTCTTCAATGCGGCTT 1019
QY	1104 CCAACTGCGGCTCCAGATCAAGGCGGCTTGACATATGCGGCAAGATCATGCGGCTGA 1163
DB	1020 CCAACTGCGGCTCCCAATCAAGGCTGCGCTTGACATATGCGGAAAGATTAATGAGCTGA 1079





Db 3204 CCAGCAGAGCTCTTCATGTGTGCCACAGTATGATATACCTCACCCCTGAACAACGGAGTCAg 3263  
 QY 3381 GCCGTGGAGCTTCATCCTTTTACTGCTGGAATATTTCCCTTCGAGATGCTGAGAAGC 3440  
 Db 3264 GCAGTAGGACGCTCTTCATTTTACTGCTGAGTACTTCCCTCTGAGATGCTGCGTACC 3323  
 QY 3441 GGCAACAACCTTACCTTCAGCTACACCTTTGAGAGAGTGCCTTTCCACAGCAGCTACGCG 3500  
 Db 3324 GGAACAACCTTACCTTCAGCTACACCTTTGAGAGAGTGCCTTTCCACAGCAGCTACGCT 3383  
 QY 3501 CACAGCCAGAGCTGGAACCGGCTGATGAATCTCTCATCGACCAATACCTGTATTACCTG 3560  
 Db 3384 CACAGCCAGAGCTGGAACCGTCTCATGAATCTCTCATCGACCAATACCTGTATTACCTG 3443  
 QY 3561 AACAGAACTCAAAATCAGTCCGGAAGTCCCAAAACAAGACTTGTGTTAGCCGTGGG 3620  
 Db 3444 AGCAGAAACAACACTCCAAGTGAACCAACGAGTCAAGGCTTCAGTTTCTCAGGCC 3503  
 QY 3621 TCTCCAGCTGGCATGTCTGTTCAAGCCCAAAATGAGCTACCTGAGCCCTGTATCGGCAG 3680  
 Db 3504 GGAGCAGTGAATTCGGGACCAAGTCTAGGAAGTGGCTTCTGAGACCTGTATACCGCCAG 3563  
 QY 3681 CAGCGCGTTTCTAAACAACAAAACAGACAACAACAAGCAATTTTACTGAGTGTGCT 3740  
 Db 3564 CAGCGAGTATCAAGACATCTGCGGATTAACAACAAGTACTCGTGAAGTGTGAGCT 3623  
 QY 3741 TCATAATATTAACCTCAATGGGCGTGAATCCATCATCAACCTGAGCATGTGCGCTCA 3800  
 Db 3624 ACCAAGTACCACTCTCAATGGCAGAGACTCTGTGTAATCCGGCCCGCATGGCAAGC 3683  
 QY 3801 CACAAAGAGCAGCAAGCAAGTCTTCCCATGACCGGTGTATGATTTTGGAAAAAGAG 3860  
 Db 3684 CACAAAGAGCAGTGAAGAAAGTTTTCCTCAGAGCGGGTCTCATCTTTGGGAAGCAA 3743  
 QY 3861 AGCGCCGAGCTTCAACAACCTGCAATGGACAATGTATGATTACAGACGAAGAGAAAT 3920  
 Db 3744 GGCTCAGAGAAACAATGTGAACATTGAAAGGTCAATGATTACAGACGAAGAGAAATC 3803  
 QY 3921 AAAGCCACTAACCTGTGGCCACCGAAAGATTTGGGACCGTGGCAGTCAATTTCCAGAGC 3980  
 Db 3804 GGAACAACCAATCCCGTGGCTACGAGCAGTATGTTCTGTATCTACCAACCTCCAGAGA 3863  
 QY 3981 AGCAGCAGAGACCTGCGACCGGAGATGTGATGCTATGGAGCATTAACCTGGCATGTG 4040  
 Db 3864 GGCAACAGACAGAGCTACCGCAGATGTCAACAACAAGCGTCTTCAGGCATGTGTC 3923  
 QY 4041 TGGCAAGATAGAGACGTGTACCTGACAGGCTCCCATTTGGGCCAAATTCCTACACAGAT 4100  
 Db 3924 TGGCAGGACAGAGATGTGTACCTTACAGGGGCCCATCTGGGCAAGATTCCACACAGGAC 3983  
 QY 4101 GGACACTTTCACCCGCTCTCTTATGGGCGGCTTTGGAATCAAGAACCCCGCTCTCAG 4160  
 Db 3984 GGACATTTTCACCCCTCTCTCTCATGGGTGATTCGAACTTAAACACCCCTCTCTCCACAG 4043  
 QY 4161 ATCTCATCAAAAAACAGCGCTGTCTCTGCGAATCTCTCGGCGAGTTTTCAGCTACAAAG 4220  
 Db 4044 ATTCTCATCAAGAAACACCCCGGTACCTGCGAATCTTTCGACCACTTCAGTGGGCAAG 4103  
 QY 4221 TTGCTTCATTCATCAACCAATATCTCCACAGACAGAGTGAATGTGAATTTGAATGGAG 4280  
 Db 4104 TTGCTTCCTTCATCAACAGTACTCCACGGGACAGCGTCAAGATGAGTGGAG 4163  
 QY 4281 CTGCAGAAAGAAAAACAGAGCGCTGGAATCCGGAAGTGCAGTACACATCAATATATGCA 4340  
 Db 4164 CTGCAGAAAGAAAAACAGAAACGCTGGAATCCGGAATTCAGTACACTTCAACTACAAC 4223  
 QY 4341 AAATCTGCCAACGTTGATTTTACTGTGACACAAATGACTTTTACTGAGCTCGCCCC 4400  
 Db 4224 AAGTCTGTATCTGTGAGCTTACCGTGATATCTAATGCGCTGTATTCAGAGGCTCGCCCC 4283  
 QY 4401. ATGGCACCCTGTTACCTTACCTGCTCCCTGTATATGAGTGTATCAATAAACCAGGTGA 4460  
 Db 4284 ATGGCACCAGATACCTGACTGTATATCTGTATGTTGTATCAATAAACCAGTTTAA 4343

QY 4461 TTGCTTCAGTGAACCTTGTGCTCTCTGCTCTTCTTATCTTACGGTTACCATGGTTATA 4520  
 Db 4344 TTGCTTCAGTGAACCTTGTGCTCTCTGCGTATTTCTTCTTATCTAGTTTCCATGCTCTA 4403  
 QY 4521 GCTTAC 4526  
 Db 4404 GACTTAC 4409  
 Db 4404 GACTTAC 4409  
 RESULT 14  
 AAV21648  
 ID AAV21648 standard; DNA; 4767 BP.  
 XX AAV21648;  
 AC AAV4 genome.  
 DT 28-AUG-1998 (first entry)  
 XX  
 DE AAV4 genome.  
 XX AAV; AAV-4; Rep gene; VP1 gene; capsid protein; vector;  
 KM gene transfer; gene delivery; cancer; gene therapy; ds.  
 XX Adeno associated virus 4.  
 OS  
 XX  
 FH Key location/Qualifiers  
 FT promoter 130..291  
 FT /\*tag= a  
 FT /note= "p5"  
 FT CDS 372..2243  
 FT /\*tag= b  
 FT /product= Rep protein  
 FT CDS 2260..4467  
 FT /\*tag= c  
 FT /product= capsid protein  
 PN MO9811244-A2.  
 XX  
 PD 19-MAR-1998.  
 XX  
 PF 11-SEP-1997; 97WO-US16266.  
 XX  
 PR 11-SEP-1996; 96US-0025934.  
 XX  
 PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 XX  
 PI Chlorini JA, Kotin RM, Safer B;  
 XX  
 DR WPI; 1998-207403/18.  
 DR P-PSDB; AAW46307-08.  
 DR  
 XX  
 XX  
 PT Adeno-associated virus 4 based vectors - used to transduce erythroid  
 PT progenitor cells for treatment of e.g. cancer and other genetic  
 PT diseases  
 XX  
 PS Claim 12; Page 46-49; 80bp; English.  
 XX  
 CC This nucleotide sequence comprises the adeno associated virus 4  
 CC (AAV4) genome that includes open reading frames for the Rep protein  
 CC (see AAW46307) and the VP1 capsid protein (see AAW46308). The genome  
 CC was cloned from viral lysates amplified from COS and HeLa cells.  
 CC The invention provides recombinant vectors and viral particles based  
 CC on AAV4 that may be useful for transducing erythroid progenitor cells  
 CC e.g. for the treatment of cancer and genetic diseases which can be  
 CC corrected by bone marrow transplants using matched donors. A  
 CC claimed method of delivering a nucleic acid to a subject (including  
 CC a subject with antibodies to AAV2) comprises administering to a  
 CC cell from the subject an AAV4 particle comprising the nucleic acid  
 CC inserted between a pair of AAV inverted terminal repeats (see  
 CC AAV21651 and AAV21659), and returning the cell to the subject. Also  
 CC claimed is a method of screening for infectivity by AAV4 comprising  
 CC contacting the cell with AAV4 and detecting the presence of AAV4 in  
 CC the cells. Cells containing the AAV4 genome, ORI, OR4 or

CC Individual Rep protein and capsid proteins genes are provided.

XX Sequence 4767 BP; 1137 A; 1404 C; 1333 G; 891 T; 2 other;

Query Match 55.0%; Score 2596; DB 19; Length 4767;  
Best Local Similarity 74.5%; Pred. No. 0;  
Matches 3582; Conservative 0; Mismatches 1096; Indels 129; Gaps 20;

QY 1 TTGCCACTCCCTCTCTGCGCGCTCGCTCGGTGGGCGCTGGCAACCAAGTCCGC 60  
DB 1 TTGGCCACTCCCTCTATGCGCGCTCGCTCACTCGGCCCTGGAGCAAAAGTCTCC 60  
QY 61 AGACGGCAGAGCTTGCTCTGCGCGCGCCCAACCGAGCGAGCGCGCAGAGGAGTG 120  
DB 61 AGACTGCGCGCGCTTGCGCGCGCGCAGCGAGGTGAGCGAGCGCGCATAGAGGAGTG 120  
QY 121 GGCAACTCCATCACTAGGG-----GTAAATCGCGAAGCGCCTCCAC 161  
DB 121 GCCAACTCCATCATCTAGGTTTGCCCACTAGCATGATGTACGCTCTAGGGTTAGGAG 180  
QY 162 GCTGCCCGCTGAGCGCTGACGTAATTAAGTATAGGAGTGTCTGTGTA-----T 213  
DB 181 GTCCCTGTATAGCAGTCAAGTGTCTGTATTTCCGAGCGTACGCGAGCGCATACC 240  
QY 214 TAGCTGTACGTGA-----GTGCTTTGCGACATTTTGCGACACCAAGTGGCAT 263  
DB 241 AAGCTGCCAGCTACAGCCACGTGTCCGTTTGCGACAGTTTGCGACACCAATGTGTAG 300  
QY 264 TTAGGATATATAGCCGAGTGAAGCAGC-AGGATCTCCATTTTGACCGGAAATTTGAA 322  
DB 301 G-AGGATATATACCGGAGTGAAGCAGCGAGAGCTCCATTTTGCGCGAAATTTGAA 359  
QY 323 CGAGCAGCAGCGCATCGCGGCTTCTACGAGTGTGATCAAGGTGCGGAGCGACCTGAC 382  
DB 360 CGAGCAGCAGCGCATCGCGGCTTCTACGAGTGTGATCAAGGTGCGGAGCGACCTGAC 419  
QY 383 GAGCACCTGCGCGGCAATTTCTGACTCGTTTGAGCTGGTGGCGGAGAAAGATGGAG 442  
DB 420 GAGCACCTGCGCGGCAATTTCTGACTCGTTTGAGCTGGTGGCGGAGAAAGATGGAG 479  
QY 443 CTGCCCCCGGATTTCTGACATGATCTGAATCTGATTGAGCAGGCAACCTGACCGTGCC 502  
DB 480 CTGCCCCCGGATTTCTGACATGATCTGAATCTGATTGAGCAGGCAACCTGACCGTGCC 539  
QY 503 GAGAGCTGACGCGCACTTCTGTCTCAATGCGCGCGGTGAGTAAGCCCCGAGGCC 562  
DB 540 GAAAGCTGACGCGCACTTCTGTCTCAATGCGCGCGGTGAGTAAGCCCCGAGGCC 599  
QY 563 CTCTTCTTTGTTCAATTGAGAAAGGCGAGTCTACTTCCACCTTCATATTCTGTGAG 622  
DB 600 CTCTTCTTTGTTCAATTGAGAAAGGCGAGTCTACTTCCACCTTCATATTCTGTGAG 659  
QY 623 ACCACGGGGGTCAAAATCCATGTGTGCGCGCTTCTGAGTCAAGTTAGGAGCAAGCTG 682  
DB 660 ACCGTGGGCGTCAAAATCCATGTGTGCGCGCTTCTGAGTCAAGTTAGGAGCAAGCTG 719  
QY 683 GTGACAGACCATCTACCGCGGATCGAGCGGACCTGCGCAACCTGTTGCGGTGACCAAG 742  
DB 720 GTGACCGGATCTACCGCGGATCGAGCGGACCTGCGCAACCTGTTGCGGTGACCAAG 779  
QY 743 ACCGCTAATGCGCGCGGAGGCGGAGCAAGGTGTGAGCAGAGTGTACTATCCCACTAC 802  
DB 780 ACCGCTAATGCGCGCGGAGGCGGAGCAAGGTGTGAGCAGAGTGTACTATCCCACTAC 839  
QY 803 CTCCTGCCCAAGACTCAGCCCGAGTGTGAGTGGCGGTGAGTGAATGAGAGTATATA 862  
DB 840 CTGCTCCCGCAAGACCCAGCGCTCCAGTGGCGGTGAGTGAATGAGAGTATATA 899  
QY 863 AGGCGGTGTGAACCTGCGCGAGCGCAAGCGCTGTGCGCAGCAGCTGACCCAGCTG 922  
DB 900 AGCGCTGTGTAATCTGCGGAGCGTAACGGCTGTGCGCAGCATCTGACGCAAGTGTG 959  
QY 923 AGCCAGACCAAGAGCAGAACAGAGAAATCTGAACCCCAATTTGACGCGCTGTATC 982

DB 960 TCGCAGACCGCAGAGCAGAGCAAGAAACAGAACCCCAATTCTGACGCGCGGTATC 1019  
QY 983 CGGTCAAAAACCTCCGCGGCTTACATGAGCTGTGCGGTGCTGTGACCGGGGATC 1042  
DB 1020 AGGTCAAAAACCTCCGCGGCTTACATGAGCTGTGCGGTGCTGTGACCGGGGATC 1079  
QY 1043 ACCTCGAAGACAGTGTGATCAGAGAGACAGGCTCTGTACATCTCTTCAACGCGCT 1102  
DB 1080 ACGTCAAAAACAGATGATCAGAGAGACAGGCTCTGTACATCTCTTCAACGCGCT 1139  
QY 1103 TCCAACTCGCGTCCAGATCAAGCGCTGTGACAAATGCGCGCAAGATCATGCGCTG 1162  
DB 1140 TCCAACTCGCGTCCAGATCAAGCGCTGTGACAAATGCGCTCAAAATCATGAGCTG 1199  
QY 1163 ACCAAATCCGCGCGCGCTACCTGTAGGCGCGCTCCGCGCGGACATTAACCAAC 1222  
DB 1200 ACAAGACGCGCTCCGAGCTACCTGTGCGCGAGAACCGCGCGGACATTTCCAGCAC 1259  
QY 1223 CGCATCTACCGCATCTCTGAGCTGAACGCTACGAACCTGCTTACGCGGCTCGCTTT 1282  
DB 1260 CGCATCTACCGCATCTCTGAGTGAACGCTACGATCCGAGTACGCGCTCGCTTT 1319  
QY 1283 CTGCGTGGCGCCAGAAAAGTTGCGGAGCGCAACCATCTGCTGTTGGCGCGCC 1342  
DB 1320 CTGCGTGGCGCCAGAAAAGTTGCGGAGCGCAACCATCTGCTGTTGGCGCGCC 1379  
QY 1343 ACCACGGGCAAGACCAACATCGCGGAAAGCATCGCCAGCGCTGCTTACGCGTGC 1402  
DB 1380 ACAGCGGGTAAACCAACATCGCGGAAAGCATCGCCAGCGCTGCTTACGCGTGC 1439  
QY 1403 GTCAATGGAACCAATGAGACTTTCCCTTCAATGATGTGCTGACAAAGATGTATCTG 1462  
DB 1440 GTGAATGGAACCAATGAGACTTTCCCTTCAACGATGTGCTGACAAAGATGTATCTG 1499  
QY 1463 TGGAGGAGGGCAAGATGACGCGCAAGGTGTGAGTCCGCCAAGGCCATTCTCGCGGC 1522  
DB 1500 TGGAGGAGGGCAAGATGACGCGCAAGGTGTGAGTCCGCCAAGGCCATTCTCGCGGC 1559  
QY 1523 AGCAAGGTGCGCGTGACCAAAAGTCAAGTGTGCGCGCAAGTCAACCCACCGCGTG 1582  
DB 1560 AGCAAGGTGCGCGTGACCAAAAGTCAAGTGTGCGCGCAAGTCAACCCACCGCGTG 1619  
QY 1583 ATGCTCACTTCAACCAACCAATGTGCGCGGTGATGACGGAACAGCAACCACTTGAG 1642  
DB 1620 ATGCTCACTTCAACCAACCAATGTGCGCGGTGATGACGGAACAGCAACCACTTGAG 1679  
QY 1643 CACAGCAGCGGCTGACAGACCGGATGTTCAATTTGAATCAACCGCGCTGTGAGCAT 1702  
DB 1680 CACAGCAGCGGCTGACAGACCGGATGTTCAATTTGAATCAACCGCGCTGTGAGCAT 1739  
QY 1703 GACTTTGGCAAGGTGACCAAGCAGGAAGTCAAAAGATTTCTCCGCTGGCGCAGATCAC 1762  
DB 1740 GACTTTGGCAAGGTGACCAAGCAGGAAGTCAAAAGATTTCTCCGCTGGCGCAGATCAC 1799  
QY 1763 GTGACCGAGGTGCGCATGATTCTACGTCAAGAAAGGTGAGCCCAAAAGAGCCGCC 1822  
DB 1800 GTGACCGAGGTGACTACGAGTTTTCGTCAAGAAAGGTGAGCTAGAAAGAGCCGCC 1859  
QY 1823 CCGCATGACGCGGATTAAGCGAGCCCAAGCGGCTGCGCTCAGTCCGCGATCATCG 1882  
DB 1860 CCGCATGACGCGGATTAAGCGAGCCCAAGCGGCTGCGCTCAGTCCGCGATCATCG 1919  
QY 1883 ACCTCAGACGCGGAGAGCTCCGCTGAGCTTTGCGCAGAGTACCAAAACAAATGTTCT 1942  
DB 1920 ACCTCAGACGCGGAGAGCTCCGCTGAGCTTTGCGCAGAGTACCAAAACAAATGTTCT 1976  
QY 1943 CGTCAAGCGGCGATGCTTCAAGTGTGTTCCCTGCAAGACATGCGAGAGATGATCAG 2002  
DB 1977 CGTCAAGTGGTATGATCTGATGCTTTTCCCTGCGGCAATGCGAGAGATGATCAG 2036  
QY 2003 AATTCAACATTTGCTTCAAGCAGCGGAGCAGAGACTGTTCAAGTGTGTTCCCGCGCTG 2062



Db 2037 AATGTGACATTGCTTCAACGACGGGGTCATGACTGTGCCAGTGTCC--CGTG 2093  
QY 2063 TCAGANTCTCAAC-----CGGTGTCAGAAAGAGACGTATCGAAACTGTGCCATT 2116  
Db 2094 TCAGATCTCAACCCCGTGTGTGTGTGTCAGAAAGCGACGTATCAGAAACTGTGCCATT 2153  
QY 2117 CATCATCTGCTGGGGGGGCTCCCGAGATTGCTTGGCTCGGCTGCGATCTGTCAACGTG 2176  
Db 2154 CATCATCATGCGGGGGGGCGCCGAGGTGGCTGTCCGCTGCGAAGTGGCCAATGTG 2213  
QY 2177 GACCTGATGACTGTGTTCTGAGCAATAATGACTTAAACAGGTATGCTGCCGATGG 2236  
Db 2214 GACTTGATGACTGTGACATGGAAACAATAATGACTCAAAACGATATGACT--GACGG 2270  
QY 2237 TTATCTTCAGATTGGCTTCGAGGACAACTCTGTGAGGGCATTCGCGAGTGTGGACTT 2296  
Db 2271 TTACCTTCAGATTGGCTTAGAGGACAACTCTGTGAAGGCGTTGAGAGTGTGGGCGCT 2330  
QY 2297 GAAACCTGAGCCCGGAGCCCAAGCCCAAGCAAAAGCAGCAAAAGCAGCGCGGGGTCT 2356  
Db 2331 GCAACCTGAGCCCTTAAACCAAGGCAATCAACAATCAGACAAACGCTCGGGGTCT 2390  
QY 2357 GGTGCTTCTGCTCAAGTACTCTGAGACCTTCAACGACTCGACAAAGGGGAGCCCGT 2416  
Db 2391 TGTGCTTCCGGTTACAATAAATCTCGAACCCGGAACGGAATCGAACAGGGGGAACCCGT 2450  
QY 2417 CAACGGCGGAGCGGAGCGGCGCTCGAGACGACGAGCGCTTCGACGAGCTCAAGC 2476  
Db 2451 CAACGAGCGGAGCGGCGGAGCGCTCGAGACGACGAGCGCTTCGACGAGCTCAAGC 2510  
QY 2477 GGGTGACAATCCGTAATCTGCGGTATTAACACGCGCGAGCTTCAGAGCGCTGCA 2536  
Db 2511 CGGTGACAACCCCTTACTCAAGTACAACACGCGCGAGTTCAGACGCGGCTTCA 2570  
QY 2537 AGAAGATACGCTTTTGGGGGGCAACCTCGGGCGAGCATCTTCAGGGCCAAGACGGGT 2596  
Db 2571 GGGCGACACACCGTTGGGGGGCAACCTCGGACAGCATCTTCAGGGCCAAGACGGGT 2630  
QY 2597 TCTGAACTCTCGGTCTGTGTGAGGAGGCGCTAAGCGGCTCTGGAAGAAAGCTCC 2656  
Db 2631 TCTTGAACCTCTGTGTGTGAGGAGGCGGTGAGCGGCTCTGGAAGAAAGAGACC 2690  
QY 2657 GGTAGAGCAGTCCGCAAGAGCCAGACTCTCTCGGGCATCGGCAAGACAGGCGACA 2716  
Db 2691 GTTGATTGAATCCCCCGAGAGCGCCGACTCTCCAGCGGTATCGCAAAAAGGCAAGCA 2750  
QY 2717 GCGCGTAAAAAGAGACTCAATTTTGTGACACTGCGACTCAGAGTCCCGCATCC 2776  
Db 2751 GCGCGCTAAAAAGAGCTCGTTT-----CGAAGACGAAACTGAGCGGCGACGG 2801  
QY 2777 ACAACCTCTCGAGAACCTTCAGCAACCCCGCTGTGTGGACCTACTACATGGCTTC 2836  
Db 2802 ACCCCCTGAGGATCACTTCGGAGCCATGTGTGATGACAG-----TGAGATGCGTGC 2855  
QY 2837 AGGCGGTGGCGCAATGCGACACAATAACGAAGCGCGCGAGGTGGTAAATGCTTC 2896  
Db 2856 AGCAGCTGGCGAGCTGCAGTGAGGGGGACAAAGTGCCTGATGAGTGGTAAATGCTTC 2915  
QY 2897 AGGAATTTGCAATGCGATTTCACATGCTGGCGCAGAGTCAATCACCACGACCCG 2956  
Db 2916 GGGTGATTGCAATGCGATTTCACCTGTGTGAGGGCCACGTCAAGACCAAGACACGAG 2975  
QY 2957 CACCTGGCGCTTGGCCACTCAATAACCACTTACAGCAATCTCCAGTGTCTCAAC 3016  
Db 2976 AACCTGGGTCTTGGCCACTCAACAACCACTTACAGCAATCTCCAGTGTCTCAAC 3024  
QY 3017 GGGGGCCAGCAACCACTACTTGGGTACAGACCCCTGGGGGTATTTGATTT 3076  
Db 3025 -GAGAGCTGCAAGTCCACAACCTAACAAGGATTTCTCAACCCCTGGGGATTTGACTT 3083  
QY 3077 CAACAGATTCCACTGCACTTTTCAACCAAGTACTGCGAGCGACTCATCAACAATTTG 3136  
Db 3084 CAACCGCTTCACTGCACTTCTCAACCAAGTACTGCGAGCGACTCATCAACAACCACTG 3143

QY 3137 GGGATTCGGCCCAAGAGACTCACTTCAAACTCTTCAACATCCAACTCAAGAGGTCA 3196  
Db 3144 GGGCATGCGAACCCCAAGCCATGCGGGTCAAAATCTTCAACATCCAGTCAAGAGGTCA 3203  
QY 3197 GACGATGATGGCGTCAACCACTCCCTAATAACCTTACCAGACAGGTTCAAGTCTTTC 3256  
Db 3204 GACGTGCAACGGCGAGACACACGGTGGCTAATAACCTTACCAGACAGGTTCAAGTCTTTC 3263  
QY 3257 GGAATCGAGTACCAAGCTTCCGTACGTCTCGGCTCTGCGCACAGGCTGCTCCCTCC 3316  
Db 3264 GGAATCGTGTACGAAGTCCGTACGTATGATGATGCGGGTCAAGAGGCGAGCTGCTCC 3323  
QY 3317 GTTCCCGGAGCGTGTTCATGATTCGCAATACGGCTAC-----CTGACGCTCAA 3367  
Db 3324 TTTTCCCAACGACGTCTTATGTGCCCCAGTACGGCTACTGTGAACTGTGACCGGCAA 3383  
QY 3368 CAATGGACCCAAAGCCGTGGAGCTTCACTCTTTACTGCTGGAATATTTCCCTTCTCA 3427  
Db 3384 CACTTGCAGCAACAGACTGACAGAAATGCTTCTACTGCTGGAATACTTCTTCCGA 3443  
QY 3428 GATGCTGAGAACGGGACAACTTTACCTTACGCTACACCTTTGAGGAAGTGCCTTCCA 3487  
Db 3444 GATGCTGCGGACTGGCAACAACTTTGAATTAAGTACAGTTTGAAGAGTGCCTTCCA 3503  
QY 3488 CAGCAGCTACCGCGACAGCCAGAGCTTGAACCGGCTGATGAATCTCTCATCGACCAATA 3547  
Db 3504 CTGATGATACCGCGACAGCCAGAGCTTGAACCGGCTGATGAATCTCTCATCGACCAATA 3563  
QY 3548 CCGTATTAATCTGAACAGAACTCAAAATCAGTCCGGA--AGTCCCCAAAACAAGACTT 3604  
Db 3564 CCTGTGGGACTGCAATGACCAACACCGGAACCACTGATATGCCGGAAGTCCACAC 3623  
QY 3605 GCTGTTAGCCGTGGGTCTCAGCTGCGATGTCTGTTACGCCCAAAAAGTGTACCTGG 3664  
Db 3624 CAACCTTACCAAGCTGCGGCTTCAAACTTTTCCAACTTTAAAGAAAGTGTGCGCG 3683  
QY 3665 ACCCTGTATCGGACAGCGCGTCTTAAACAAACAAACAAACAAACAAACAACTTT 3724  
Db 3684 GCCTTCAATCAAGCAGAGGCTCTCAAGAGCTGCCAATCAAAAGTATCCCTGC 3743  
QY 3725 TACCTGACTGTGCTTCAAAATATAAC-----CTCAATGGGCGTGAATC 3769  
Db 3744 CACCGGCTCAGACAGTCTCATTAATAAGAGACGACAGCACTGTGAACGGAAGATGAG 3803  
QY 3770 CATCATCAACCTTGGCACTGTGTATGCTTCAACAAAGACGACGAAGCAAGTCTTTC 3829  
Db 3804 TGCCCTGACCCCGGACCTCAATGGCCACGCGTGAACCTGCGGAACAGCAAGTTAG-- 3860  
QY 3830 CATGAGCGGTGCTGATTTTGGAAAAAGAGAGCGCGGAGCTTCAAAACACTGCATTGA 3889  
Db 3861 CAACAGCCAGCTCATCTTTGCGGGCTTAAACAGAACGGAACAACGCGTACCGCG 3920  
QY 3890 CAATGTGATGATTACAGAGGAAGAAATTAAGCCACTAACCTGTGGCCACCGAAG 3949  
Db 3921 GACTGTGATCTTCACTCTGAGAGGAGCTGGCAGCCACCAACGCAACCGTACGACAT 3980  
QY 3950 ATTGGACCGGTGCGAGTCAATTTCCAGAGCAGCAGCAAGCCCTGCGACCGAGATGT 4009  
Db 3981 GTGGGGCAACCTACTGCGGCTGACAGCAACAGCAACCTGCGGACCGTGAACAGAT 4040  
QY 4010 GCATGCTATGGAGCATTAACCTGGCATGTGTGGCAAGATGAGACGTGTACTGCAGGG 4069  
Db 4041 GACAGCTTGGGAGCGGCTGCTGAATGTCTGGCAAAACAGAGACATTTACTACAGGG 4100  
QY 4070 TCCCATTTGGGCCAAAATTCCTACACAGATGAGACATTTCAACCGCTCTCTTATGG 4129  
Db 4101 TCCCATTTGGGCCAAGATTCCTACACAGATGAGACATTTCAACCGCTGATTTGG 4160  
QY 4130 CGGCTTGGACTCAAGAACCGGCTCTCAGATCCTCATCAAAAAACAGCGCTGTCTGC 4189  
Db 4161 TGGGTTGGGCTGAACAACCGGCTCTCAAAATTTTATCAAGAAACAACCGGTACTCTGC 4220

QY	4190	GAATCCTCCGGCGGAGTTTTCAGCTACAAAGTTTGCTTCATTCAATCACCACCAATATCCAC	4249
Db	4221	GAATCCTGCAACGACCTTCAGCTCTACTCCGTTAACTCTTCAATTACTCAGTACAGAC	4280
QY	4250	AGGCAAGTAGTGTGGAAATTGAATGGAGCTGCAGAAAGAAAACAGCAAGCGCTGAA	4309
Db	4281	TGGCCAGGTGTGCGTGCAAGTTGACTGGAGATCCAGAAGAGCGGTCCAAACGCTGAA	4340
QY	4310	TCCCCAAGTGCAGTACACATCCCAATTATGCAAAATCTGCCAAGTTGATTTTACTGTGA	4369
Db	4341	CCCCGAGGTCAGTTTACTCTCCAACCTACGACAGCAAAAACCTCTGTGTGGGCTCCCGA	4400
QY	4370	CAACAATGACTTTATACTGAGCCTCGCCCCATTGGCAACCGTTACCTTACCCTCCCT	4429
Db	4401	TGCGGCTGGAAATACACTGAGCCTAGGGCTATCGTTACCCGCTACCTCACCCACCACT	4460
QY	4430	GTAATTACGTGTTAATCAATAAACCGGTTGATTGTTTCAGTTGAACCTTGTCTCC-TG	4488
Db	4461	GTAATAACCTGTTAATCAATAAACCGGTTTATTTCGTTTCAGTTGAACCTTGTCTCCGTG	4520
QY	4489	TCCTTCTTATCTTATC-GGTTACCATGCTTATAGCTTACACATTAAGTG-----CTTGG	4541
Db	4521	TCCTTCTTATCTTATCTGCTTTCATAGGCTACTGCGTACATAAGCAGCGGCTGCGGCGC	4580
QY	4542	TTGCGCTTCGCGATAAAGACTTACG-----TCATCGGGTTACCCCTAGTGATGG	4591
Db	4581	TTGCGCTTCGCGTTTACAACCTGCGGTTAATCAGTAACTTCTGGCAAAACAGATGATGG	4640
QY	4592	AGTTGCCCACTCCCTCTCTGCGCGCTCGCTCGCTCGGTGGGGCTCGCGAACCAAGTCC	4651
Db	4641	AGTTGGCCACATTAAGTATGCGCGCTCGCTCACTCACTCGGCTTGAGAGACCAAGTCT	4700
QY	4652	GCAGACGCGAGAGCTCTGCTCTGCCGGCCCAACCGAGCGAGCGGCGCAGAGAGGAG	4711
Db	4701	CCAGACTGCCGCGCTCTGGCCGGCAGGGCCGAGTGAGTGAGCGAGCGGCATAGAGGAG	4760
QY	4712	TGGGCCAA 4718	
Db	4761	TGGCCAA 4767	

RESULT 15	
AAD00831/c	
ID	AAD00831 standard; DNA; 7215 BP.
XX	
AC	AAD00831;
XX	
DT	08-SEP-2000 (first entry)
XX	
DE	Hybrid adeno-associated virus 2/4 helper plasmid sequence.
XX	
KW	Hybrid virus plasmid; adeno-associated virus; AAV2; AAV4; gene therapy;
KW	recombinant parvoviral vector; cellular tropism; rep gene; cap gene;
KW	capsid; gene delivery; ss.
XX	
OS	Chimeric - Adeno-associated virus serotype 2.
OS	Chimeric - Adeno-associated virus serotype 4.
XX	
FH	
Key	Location/Qualifiers
CDS	251..2183
FT	/*tag= a
FT	/product= "Rep 68"
FT	/note= "The specification states that this region
FT	encodes Rep 68 protein of AAV2. However, the positions
FT	given in the specification for this CDS do not correspond
FT	to start or stop codons"
FT	251..2120
CDS	
FT	/*tag= b
FT	/product= "Rep 78"
FT	/note= "The specification states that this region
FT	encodes Rep 78 protein of AAV2. However, the positions
FT	given in the specification for this CDS do not correspond
FT	to start or stop codons"

FT	923..2120
FT	/*tag= c
FT	/product= "Rep 52"
FT	/note= "The specification states that this region
FT	encodes Rep 52 protein of AAV2. However, the positions
FT	given in the specification for this CDS do not correspond
FT	to start or stop codons"
FT	923..2183
FT	/*tag= d
FT	/product= "Rep 40"
FT	/note= "The specification states that this region
FT	encodes Rep 40 protein of AAV2. However, the positions
FT	given in the specification for this CDS do not correspond
FT	to start or stop codons"
FT	2123..4341
FT	/*tag= e
FT	/label= "Cap protein region"
FT	/note= "The Cap protein region contains the coding
FT	regions for the proteins Vp1, Vp2 and Vp3 where the
FT	start positions for these CDS are 2123, 2547 and 2727
FT	respectively. However, the positions given in the
FT	specification for this CDS do not correspond to start
FT	or stop codons"

PN WO200028004-A1.  
XX  
PD 18-MAY-2000.  
XX  
PF 10-NOV-1999; 99WO-US26505.  
XX  
PR 10-NOV-1998; 98US-0107840.  
PR 10-MAR-1999; 99US-0123651.  
XX  
XX  
PA (UYNC-) UNIV NORTH CAROLINA.  
XX  
PI Rabinowitz JE, Samulski RJ, Xiao W;  
XX  
DR WPI; 2000-376523/32.

PT Recombinant parvoviral vectors with altered packaging, tropisms and  
immunogenic properties, useful in gene therapy protocols -  
XX  
PS Example 28; Page 130-134; 153pp; English.

245 The patent discloses modified parvovirus vectors with advantageous  
246 antigenic properties, packaging capabilities and cellular tropisms.  
247 These vectors can be used in standard recombinant DNA protocols e.g. gene  
248 therapy for delivering nucleic acids to cells. The present sequence is  
249 a hybrid virus plasmid comprising adeno-associated virus serotype 2  
250 (AAV2) rep genes and AAV4 capsid (cap) genes in a pBluescript backbone.  
251 The rep gene encodes four proteins responsible for replication,  
252 Rep 68, Rep 78, Rep 52 and Rep 40 and the cap gene encodes three  
253 structural proteins VP1, VP2 and VP3. The hybrid virus is used to  
254 introduce a heterologous gene into a target cell.

Sequence 7215 BP; 1548 A; 1918 C; 1925 G; 1824 T; 0 other;

Query Match	51.7%;	Score 2438.8;	DB 21;	Length 7215;
Best Local Similarity	75.0%;	Pred. No. 0;		
Matches 3273; Conservative	0;	Mismatches 1012;	Indels 77;	Gaps 15;

[illegible]

QY	384	AGC	ACTG	CGCGGG	CATT	TTCT	GACT	CGTTT	TGTG	AGCT	GGGTG	GC	GAG	AAG	GAAT	GGAG	C	443																						
Db	4823	GGC	ATCT	GCCCGG	CATT	TTCT	GAC	AGCT	TTGT	GA	CTGGT	GC	GAG	AAG	GAAT	GGAG	T	4764																						
QY	444	TG	CCCCGG	AT	CTG	ACAT	GAT	CTGA	AT	CTGA	TTG	AG	CAG	GC	AC	CGT	GG	503																						
Db	4763	TG	CCCG	CAG	AT	CTG	ACAT	GAT	CTGA	AT	CTGA	TTG	AG	CAG	GC	AC	CGT	GG	4704																					
QY	504	AGA	AG	TC	GAC	GC	GC	AG	CTT	CTG	TG	CCA	TG	GC	CG	CG	TGA	TAA	GG	563																				
Db	4703	AGA	AG	TC	GAC	GC	GC	AG	CTT	CTG	TG	CCA	TG	GC	CG	CG	TGA	TAA	GG	4644																				
QY	564	TCT	CTTT	GT	TC	AG	TT	CG	AG	AG	GC	AG	TC	CT	AC	CT	CC	AT	AT	TT	CT	GT	GG	AGA	623															
Db	4643	TTT	CTTT	GT	TC	AG	TT	CG	AG	AG	GC	AG	TC	CT	AC	CT	CC	AT	AT	TT	CT	GT	GG	AGA	4584															
QY	624	CC	AC	GG	GG	GT	CA	AA	T	CC	AT	G	T	G	T	G	C	C	CT	T	C	T	G	A	G	CT	G	683												
Db	4583	CC	AC	GG	GG	GT	CA	AA	T	CC	AT	G	T	G	T	G	C	C	CT	T	C	T	G	A	G	CT	G	4524												
QY	684	TG	C	A	G	A	C	C	A	T	C	T	A	C	C	G	G	A	T	C	A	G	C	C	G	A	T	C	743											
Db	4523	TT	C	A	G	A	A	T	T	A	C	C	G	G	A	T	C	A	G	C	C	G	A	T	C	A	G	C	4464											
QY	744	CG	C	G	T	A	T	G	G	C	G	C	G	G	G	G	A	C	A	A	G	T	G	T	G	A	C	A	G	803										
Db	4463	CC	A	G	A	A	T	G	G	C	G	C	G	G	G	A	C	A	A	G	T	G	T	G	A	C	A	G	4404											
QY	804	T	C	T	G	C	C	A	A	G	A	C	T	C	A	G	C	C	G	A	G	T	G	G	C	G	T	G	A	C	T	A	A	863						
Db	4403	T	G	C	T	G	C	C	A	A	G	A	C	T	C	A	G	C	C	G	A	G	T	G	G	C	G	T	G	A	C	T	A	A	4344					
QY	864	G	G	C	C	T	G	T	T	G	A	C	T	G	G	C	C	G	A	C	C	G	C	T	G	T	G	G	C	A	C	C	T	G	A	C	923			
Db	4343	G	G	C	C	T	G	T	T	G	A	C	T	G	G	C	C	G	A	C	C	G	C	T	G	T	G	G	C	A	C	C	T	G	A	C	4284			
QY	924	G	C	C	A	G	A	C	C	C	A	G	A	C	A	A	G	A	A	T	C	T	G	A	A	C	C	C	C	A	A	T	T	G	A	C	983			
Db	4283	C	G	C	A	G	A	C	C	C	A	G	A	C	A	A	G	A	A	T	C	T	G	A	A	C	C	C	C	A	A	T	T	G	A	C	4224			
QY	984	G	G	T	C	A	A	A	A	C	C	T	C	G	C	G	C	T	C	A	T	G	A	G	C	T	G	T	G	T	G	A	C	C	G	G	C	A	1043	
Db	4223	G	A	T	C	A	A	A	A	C	C	T	C	G	C	G	C	T	C	A	T	G	A	G	C	T	G	T	G	T	G	A	C	C	G	G	C	A	4164	
QY	1044	C	C	T	C	G	A	A	A	G	C	A	G	T	G	A	T	C	A	G	A	G	A	C	C	A	G	C	C	T	C	T	A	C	A	C	G	C	T	1103
Db	4163	C	C	T	C	G	A	A	A	G	C	A	G	T	G	A	T	C	A	G	A																			

Db	3743	GGAGAGAGGGGAAGATGACCGCCAAAGTCTGTGAGTCCGGCCAAAGCCATTCTCGAGAGAA	3684
QY	1524	GCAAGGTGCGGTGAGACCAAAAGTCCAAGTCTGCCGCCAGATCGACCCCAACCCCGCTGA	1583
Db	3683	GCAAGGTGCGGTGAGACCAAAAGTCCAAGTCTGCCGCCAGATAGACCCGACTCCCGTGA	3624
QY	1584	TGCTCACTCCAAACACCAATCATGTGCGCGCTGATTGACGGGAACAGACCACTTCGAGC	1643
Db	3623	TGCTCACTCCAAACACCAATCATGTGCGCGCTGATTGACGGGAACCTCAACGACCTTCGAA	3564
QY	1644	ACCAGCAGCCGTTGACGAGACCGGATGTTCAAAATTTGAATCAACCCGCGCTGAGCATG	1703
Db	3563	ACCAGCAGCCGTTGACGAGACCGGATGTTCAAAATTTGAATCAACCCGCGCTGAGCATG	3504
QY	1704	ACTTTGGCAAGGTGACAAAGCAGGAAGTCAAAAGATTCTTCCGTGGCGCAGATCAAG	1763
Db	3503	ACTTTGGCAAGGTGACAAAGCAGGAAGTCAAAAGATTCTTCCGTGGCGCAGATCAAG	3444
QY	1764	TGACCGAGGTGCGCATGAGTTCTACGTCAGAAAGGTGAGCCAAACAAAGACCCGCC	1823
Db	3443	TGCTTGAAGGTGAGCATGATTTCTACGTCAAAAAAGGTGAGCCAAAGAAAGACCCGCC	3384
QY	1824	CCGATGACCGCGGATAAAGCCGAGCCCAAGCGGCGTCCCTCAGTCGCGGATCCATCGA	1883
Db	3383	CCAGTGACGACAGATATAGTGAGCCCAACCGGCTGCGGAGTCACTTGGCAGCCATCGA	3324
QY	1884	CGTCAGACGCGGAAGAGAGCTCCGGTGACTTTGCCGACAGGTACCAAAACAATGTTCTC	1943
Db	3323	CGTCAGACGCGGAAGAGAGCTCCGGTGACTTTGCCGACAGGTACCAAAACAATGTTCTC	3267
QY	1944	GTCACGCGGCGATGCTTCAAGATGCTGTTCCCTGCAAGACATGCGAGAAATGAATCAGA	2003
Db	3266	GTCACGCGGCGATGATTAATCTGATGCTTTTCCCTGCCGCAATGCGAGAAATGAATCAGA	3207
QY	2004	ATTTCAACATTGCTTCAACGACGCGGACGAGAGACTGTTCAAGTGTCTTCCCGCGCTGT	2063
Db	3205	ATGTGACATTTGCTTCAACGACGCGGATCATGAGTGTGCCAGTGTCTCC--CGTGT	3150
QY	2064	CAGATCTCAAC-----CGGTGTCAGAAAGAGACGTATCGGAACCTCTGTGCCATTTC	2117
Db	3149	CAGATCTCAACCCCGTGTCTGCTGTCAGAAAGCGGACGTATCAGAAACTGTGTCCGATTTC	3090
QY	2118	ATCATCTGCTGGGCGCGGCTCCCGAGATTGCTTGTCTGCGCTGCGATCTGCTCAACGTGG	2177
Db	3089	ATCATCATCTGGGCGCGGCGCGCGCGGAGTGGCTGTCTGCGCTGCGAACTGGCCCAATGTGG	3030
QY	2178	ACCTGATGACTGTGTTTCTGAGCAATTAATGACTTAAACCAAGTATGGCTGCCGATGTG	2237
Db	3029	ACTTGATGACTGTGACATGGAACAATTAATGACTCAAAACAGATATGACT--GACGGT	2973
QY	2238	TATCTTCCAGATTGCTCGAGGACCAACCTCTTGAAGGCAATTGCGAGTGTGGACTTG	2297
Db	2972	TACCTTCCAGATTGCTAGAGGACCAACCTCTTGAAGGCGTTGAGAGTGTGGCGCTG	2913
QY	2298	AAACCTGAGCCCGGAAGCCCAAGCCAAACGACGAAGCAAGCAGAGCAGCGCGGCGTCTG	2357
Db	2912	CAACCTGAGCCCGCTAAACCAAGCAATCAACATCAGGACCAACGCTCGGCGTCTT	2853
QY	2358	GTGCTTCCGCTACAGTACCTCGGACCCCTTCAACGGAATCGACCAAGGGGAGCCCGTTC	2417
Db	2852	GTGCTTCCGCTTACAAATACCTCGGACCCCGGCAACGGAATCGACCAAGGGGAGCCCGTTC	2793
QY	2418	AAAGCGGCGGACGAGCGGCGCTCGAGCAGCAGCAAGGCTTACGACCAAGCAGCTCAAAAGCG	2477
Db	2792	AAAGCAGCGGAGCGGCGAGCCCTCGAGCAGCAGCAAGGCTTACGACCAAGCAGCTCAAGGCGC	2733
QY	2478	GGTGACAATCCGTAACCTGCGGTATAACCAAGCCGACCGCGAGTTTCAAGAGCGTTCGAA	2537
Db	2732	GGTGACAACCCCTACCTCAAGTACAAACCAAGCCGACCGAGTTTCAAGCAGCGCTTCAG	2673
QY	2538	GAAATACGTTTGGGGGCAACCTCGGCGAGCAGTCTTCAGGCCAAGAGCGGCTT	2597



Db 2672 GCGCACATCGTTTGGGGCAACCTCGGCAGAGCAGTCTTCCAGGCCCAAAAAGAGGTT 2613  
QY 2598 CTGGAACCTCTCGGTCTGGTTGAGGAAGCGCTAAGACGGCTCTGGAAGAAACGTCG 2657  
Db 2612 CTGGAACCTCTGGTCTGGTGAAGCGGGTGAGACGGCTCTGGAAGAAAGAGACCG 2553  
QY 2658 GTAGAGAGTGCACCAAGAGCCAGACTCTCTCGGGCATCGCAAGACAGCCAGCAG 2717  
Db 2552 TTGATTGAATCCCCCAGCAGCCGACTCTCCACGGGTATCGCAAAAAAGCAAGCAG 2493  
QY 2718 CCGCTAAAAAGAGACTCAATTTTGTGAGACTGGGACTCAGAGTCAGTCCCGATCA 2777  
Db 2492 CCGGCTAAAAAGAGCTCGTTT-----CGAAGACGAACCTGGAAGCAGCGACGGA 2442  
QY 2778 CAACCTCTCGAGAACCTCCAGCAACCCCGCTGCTGTGGACCTACTACATGCTTCA 2837  
Db 2441 CCCCCTGAGGGATCAACTTCCGAGCCATGTCTGATG-----ACAGTGAATGCGTCA 2388  
QY 2838 GCGGCTGGCGCACCAATGGCAGACATTAACGAAGCGCCGACGAGTGGGTAATGCTCA 2897  
Db 2387 GCAGCTGGCGGAGCTGCAGTCCAGGGCGGACAAAGTGCAGATGAGTGGTAATGCTCG 2328  
QY 2898 GGAATTTGCAATTGCGATTCCACATGCTGGGCGACAGATCATCAACCAGCACCCGC 2957  
Db 2327 GGTGATTGGCATTTGCGATTCCACCTGTCTGAGGGCCACGTCAAGCACCAACAGCAG 2268  
QY 2958 ACCTGGCTTGCCACCTACAATAACCACTCTACAAGCAAACTCCAGTGTCAAG 3017  
Db 2267 ACCTGGCTTGCCACCTACAATAACCACTCTACAAGCAAACTCCAGTGTCAAG 2220  
QY 3018 GGGGCGACGACGACCAACCACTCTGCGCTACAGCACCCCGTGGGGTATTTGATTT 3077  
Db 2219 GAGAGCTGCAAGTCCCAACACCTTCAACGGAATTTCTCCACCCCTGGGGATCTTGA 2160  
QY 3078 AACAGATTCCACTGCGACTTTTCAACGAGTGAAGTGGCAGCGACTCATCAACAATTTG 3137  
Db 2159 AACCGCTTCCACTGCGACTTTTCAACGAGTGAAGTGGCAGCGACTCATCAACAATTTG 2100  
QY 3138 GGATTCGGGCGCCCAAGAGACTCAACTTCAAACTCTCAACATCAAGTCAAGAGTCAAG 3197  
Db 2099 GGCATGCGACCCAAAGCCATGCGGTCAAAATCTTCAACATCAAGTCAAGAGTCAAG 2040  
QY 3198 ACCAATGATGCGTCAACCAACATGCTAATACTTACCAGCAGGTTCAAGTCTTCTCG 3257  
Db 2039 ACCTGCAACGCGAGAGACAGGTGGCTAATACTTACCAGCAGGTTCAAGTCTTCTCG 1980  
QY 3258 GACTCGAGTACCAAGTCTCCGTAAGTCTCGGCTCTGCGCACAGGGCTGCTCCCTCG 3317  
Db 1979 GACTCGTCTGACGAAGTCTCCGTAAGTGAAGTGAAGGCGACGCTGCTCTCT 1920  
QY 3318 TTCCCGGCGGAGCTGTTCATGATTTCCGCAATACGGCTAC-----CTGACGCTCAAC 3368  
Db 1919 TTTCCTCAACGAGCTTTATGTTGCCAGTAAGGCTACTGTGAGTGAAGTGAAGGCGCAAC 1860  
QY 3369 AATGCGACGCAAGCGCTGGAGCTTCATCTTTACTGCTGGAATATTTCCCTTCTCAG 3428  
Db 1859 ACTTCGAGCAACAGACTGACAGAAATGCTTCTACTGCTGAGTACTTTCTTCTGAG 1800  
QY 3429 ATGCTGAGAACGGCAACAACCTTACCTTACGCTACACCTTTGAGGAAGTGCCTTCTCAG 3488  
Db 1799 ATGCTGCGGACTGGCAACAACCTTGAATAATACGTAAGTTTGAAGAGTGCCTTCTCAG 1740  
QY 3489 AGCAGCTACGCGCACAGCCAGAGCTGAGCCGGCTGATGAATCTCTCATGACCAATAC 3548  
Db 1739 TCGATGTAAGCGCACAGCCAGAGCTGAGCCGGCTGATGAATCCCTCTCATGACCAATAC 1680  
QY 3549 CTGTATTACTGAA-----CAGAACTCAAAATCAGTCCGGAAGTGCCTCAAAACAAGGACTT 3604  
Db 1679 CTGTGGGAGTGCATGCAATGCAACCAACCGGAACCACTGAAATGCGGAGTGCACCAACC 1620  
QY 3605 GCTGTTTACGCTGGGTCTCCAGCTGGCATGTCTGTTACGCCCAAAAAGTGGCTACCTGG 3664  
Db 1619 AACTTTACCAAGCTGCGGCTTACCAACTTTTCAACCTTTAAAGAACTGGCTGCGCGG 1560

QY 3665 ACCCTGTTATCGGACAGCGGTTTCTAAAACAAAAACAGACAACAACAGCAATTT 3724  
Db 1559 GCCTTCAATCAAGCAGAGGCTTCTCAAGACTGCGCAATCAAACTACAGATCCCTGC 1500  
QY 3725 TACTGCACTGTGCTTCAAAATATAAC-----CTCAATGGCGCTGAATC 3769  
Db 1499 CACCGGCTGACAGAGTCTCATTAATAACGAGACGACAGCACTCTGGAAGAGATGAG 1440  
QY 3770 CATCATCAACCTGCGACTGTATGCGCTCACAACAAGACGACGAAGCAAGTCTTCC 3829  
Db 1439 TGCCCTGAACCCCGGACCTTCAATGGCCACGGCTGAGCTGCGGACAGCAAGTTC--AG 1383  
QY 3830 CATGACGCGTGTCAATGATTTTGAAGAAAGAGAGCGCCGAGCTTCAAACTGCAATTGA 3889  
Db 1382 CAACAGCCAGCTCATCTTTGGCGGCGCTTAAACAGAACGCAACAGGCCACCGTACCGG 1323  
QY 3890 CAATGTCATGATTACAGACGAGAAATTAAGCCACTAACCTGTGGCCACCGAAG 3949  
Db 1322 GACTCTGATCTTCACTCTGAGAGAGAGTGGCAGCCACCAACGACCGATACGAGAT 1263  
QY 3950 ATTGGGACCGTGGCAGTCAATTTCCAGAGCAGACAGACCCCTGCGACCGAGATGT 4009  
Db 1262 GTGGGCGCAACCTACCTGGCGGTGACCAAGACAGCAACCTGCCGACCGTGAAGACT 1203  
QY 4010 GCATGCTATGGAGCATTAATCTGCGATGCTGTGGCAAGATGAGACGTGTACTGACGG 4069  
Db 1202 GACAGCCTTGGAGCGGCTGCTGAATGCTCTGGCAAAACAGAGCAATTTACTACAGGG 1143  
QY 4070 TCCCATTTGGGCCAAATTTCTCAACAGATGAGACATTTCAACCGCTCTCTTATGGG 4129  
Db 1142 TCCCATTTGGGCCAAATTTCTCAACAGATGAGACATTTCAACCGCTCTGATTTGG 1083  
QY 4130 CCGCTTTGGAATCAAGAACCGGCTCTCAGATCTCTCAAAAAACAGCGCTGTCTGC 4189  
Db 1082 TGGGTTTGGGCTGAACAACCGGCTCTCAAAATTTTATCAAGAAACACCGGTACTGCG 1023  
QY 4190 GAATCTCCGCGGAGTTTTCAGCTACAAGTTTGTCTTCAATCATCAACCAATACTCGAC 4249  
Db 1022 GAATCTGCAACGACCTTCAAGCTCTACTCCGTAACCTCTCATTAAGTACAGACGAC 963  
QY 4250 AGCAAGTGAAGTGTGAATTAAGTGGAGCTGCAGAAAGAAACAGCAAGCGCTGAA 4309  
Db 962 TGGCGAGGTGCGGTGAGATTAAGTGGAGATCCAGAAAGAGCGGTCCAAACGCTGAA 903  
QY 4310 TCCCGAAGTCAAGTACATCAATTAATGCAAAATCTGCCAAGTGAATTTACTGTGA 4369  
Db 902 CCCCAGGTCCAGTTTACTCTCAACTACGAGACAGCAAAATCTCTGTTGGGCTCCGGA 843  
QY 4370 CAACATGGAATTTATAGTGAAGCTTCCGCTGAGTGGCAGCCGTTACTTACCCGCTCCCT 4429  
Db 842 TGGCGCTGGAATAACACTGAGCTTACGCTATCGGTACCGCTACTCACCCACCACT 783  
QY 4430 GTAATTAAGTGTAAATCAATAAAGGTTGATTCGTTTGAATGAATTTGTCTCC-TG 4488  
Db 782 GTAATACTGTTAATCAATAAAGGTTTATTCGTTTGAATGAATTTGTCTCCGTTG 723  
QY 4489 TCCCTTATCTTATC-GGTTACCATGTTATAGCTTACACA 4529  
Db 722 TCCCTTATCTTATCTGTTTCAATGCGTACTGCGTACTA 681

Search completed: July 20, 2003, 09:39:47  
Job time : 1083.76 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2003, 13:19:56 ; Search time 1013.56 Seconds  
(without alignments)  
9603.029 Million cell updates/sec

Title: US-09-807-802A-1  
Perfect score: 4718  
Sequence: 1 ttgccacctccctctctgcg.....cgacagagggagtgagca 4718

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1439767 seqs, 1031500376 residues

Total number of hits satisfying chosen parameters: 2879534

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA:\*

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq1:\*  
11: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq2:\*  
12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq3:\*  
13: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*  
15: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
16: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
17: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3107.8	65.9	4679	10	US-09-804-898-1 Sequence 1, Appli
2	3107.8	65.9	4679	10	US-09-945-681-10 Sequence 10, Appli
3	3107.8	65.9	4679	15	US-10-038-972A-12 Sequence 12, Appli
4	3082.4	65.3	4680	15	US-10-077-294-1 Sequence 1, Appli
5	3082.4	65.3	4680	15	US-10-163-886-1 Sequence 1, Appli
6	3082.4	65.3	4680	15	US-10-263-127-1 Sequence 1, Appli
7	3055.8	64.8	4675	11	US-09-782-378A-1 Sequence 1, Appli
8	3055.8	64.8	4675	11	US-09-782-378A-2 Sequence 2, Appli
9	3055.8	64.8	4675	15	US-10-240-198-1 Sequence 1, Appli
10	2941.8	62.4	8179	15	US-10-205-942-5 Sequence 5, Appli
11	2730	57.9	8151	15	US-10-205-942-2 Sequence 2, Appli
12	2446.2	51.8	7214	15	US-10-205-942-1 Sequence 1, Appli
13	1872	39.7	1872	11	US-09-792-630-12 Sequence 12, Appli
14	1872	39.7	1872	12	US-09-953-351-12 Sequence 12, Appli
15	1872	39.7	1872	15	US-10-080-376-12 Sequence 12, Appli
16	1872	39.7	1872	15	US-10-082-671-18 Sequence 18, Appli

17	1872	39.7	1872	15	US-10-097-100-12	Sequence 12, Appli
18	1872	39.7	1872	15	US-10-023-208-12	Sequence 12, Appli
19	1814.4	38.5	1872	11	US-09-792-630-14	Sequence 14, Appli
20	1814.4	38.5	1872	12	US-09-953-351-14	Sequence 14, Appli
21	1814.4	38.5	1872	15	US-10-080-376-14	Sequence 14, Appli
22	1814.4	38.5	1872	15	US-10-082-671-20	Sequence 20, Appli
23	1814.4	38.5	1872	15	US-10-097-100-14	Sequence 14, Appli
24	1814.4	38.5	1872	15	US-10-023-208-14	Sequence 14, Appli
25	1451.2	30.8	1875	11	US-09-792-630-8	Sequence 8, Appli
26	1451.2	30.8	1875	12	US-09-953-351-8	Sequence 8, Appli
27	1451.2	30.8	1875	15	US-10-080-376-8	Sequence 8, Appli
28	1451.2	30.8	1875	15	US-10-082-671-14	Sequence 14, Appli
29	1451.2	30.8	1875	15	US-10-097-100-8	Sequence 8, Appli
30	1451.2	30.8	1875	15	US-10-023-208-8	Sequence 8, Appli
31	1446.4	30.7	1875	11	US-09-792-630-10	Sequence 10, Appli
32	1446.4	30.7	1875	12	US-09-953-351-10	Sequence 10, Appli
33	1446.4	30.7	1875	15	US-10-080-376-10	Sequence 10, Appli
34	1446.4	30.7	1875	15	US-10-082-671-16	Sequence 16, Appli
35	1446.4	30.7	1875	15	US-10-097-100-10	Sequence 10, Appli
36	1446.4	30.7	1875	15	US-10-023-208-10	Sequence 10, Appli
37	1444.8	30.6	1872	11	US-09-792-630-6	Sequence 6, Appli
38	1444.8	30.6	1872	12	US-09-953-351-6	Sequence 6, Appli
39	1444.8	30.6	1872	15	US-10-080-376-6	Sequence 6, Appli
40	1444.8	30.6	1872	15	US-10-082-671-12	Sequence 12, Appli
41	1444.8	30.6	1872	15	US-10-097-100-6	Sequence 6, Appli
42	1444.8	30.6	1872	15	US-10-023-208-6	Sequence 6, Appli
43	1405	29.8	1932	14	US-10-022-390-587	Sequence 587, App
44	1405	29.8	1932	14	US-10-022-390-605	Sequence 605, App
45	1405	29.8	1932	14	US-10-022-390-637	Sequence 637, App

ALIGNMENTS

RESULT 1  
US-09-804-898-1  
Sequence 1, Application US/09804898  
Patent No. US20020045264A1  
GENERAL INFORMATION:  
APPLICANT: DURING, MATTHEW  
XIAO, WEIDONG  
TITLE OF INVENTION: PRODUCTION OF CHIMERIC CAPSID VECTORS  
FILE REFERENCE: 102182-14  
CURRENT APPLICATION NUMBER: US/09/804, 898  
PRIOR FILING DATE: 2001-03-13  
PRIOR APPLICATION NUMBER: 60/189,110  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 1  
LENGTH: 4679  
TYPE: DNA  
ORGANISM: adeno-associated virus 2  
US-09-804-898-1

Query Match	65.9%;	Score 3107.8;	DB 10;	Length 4679;
Best Local Similarity	80.2%;	Pred. No. 0;		
Matches 3790;	Conservative 0;	Mismatches 882;	Indels 53;	Gaps 10;
QY	1	TTGCCCACTCCCTCTCTGCGGCTCGCTCGTGGGGCTGCGGACCAAGCTCCG	60	
Db	1	TTGGCCACTCCCTCTCTGCGGCTCGCTCGTCACTAGAGCGCGGACCAAGCTCCG	60	
QY	61	AGACGCGAGAGCTCTGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	120	
Db	61	CGACGCGCGGCTTTGCGGCGGCGGCGGCTCACTAGAGCGGCGGCGGCGGCGG	120	
QY	121	GGCACTCCATCACTAGAGGGTAATCGGAGCGGCTCCACGCTGCCGCTCAGCGTGA	180	
Db	121	GCCCACTCCATCACTAGAGGGT-----TCCTGAGGGGTGAGTCTGTA	163	
QY	181	CGTAATTACGTATAGG---GAGTGCTCTGTATTAGCTGTACAGTGAAGTCTTTTC	237	



Db 164 CGTGAATTACGTATAGGGTTAGGAGGTCTGTATATTAGAGTCAAGTGAAGTG-TTTTGC 222  
QY 238 GACATTTTGGACACCAAGTGGCCATTAGGGTATATATGCGGAGTGACGAGAGAT 297  
Db 223 GACATTTTGGACACCAATGTGTGTCACGCTGGGTATTAAAGCCGAGTGACGAGAGT 282  
QY 298 CTCCATTTTG-AACCGGAAATTGGAACGAGCAGCAGCCATGCGGGCTTCTACGAGATCG 356  
Db 283 CTCCATTTTGAAGCGGAGGTTGAAACGCGCAGCCGCAATGCGGGGTTTAAAGAGATTG 342  
QY 357 TGATCAAGGTGCGGAGCGACCTGGAAGAGCAGCTGCGGGCATTTTGTGACTGCTTGTGA 416  
Db 343 TGATTAAAGTCCCGAGCGACCTTGACGAGCATCTGCCGCAATTTCTGACAGCTTTGTGA 402  
QY 417 GCTGGGTGGCGGAGAGGAATGGAGCTGCCCCGGATTCTGACATGATCTGAATCTGA 476  
Db 403 ACTGGGTGGCGGAGAGGAATGGAGTGGCCGAGATTCTGACATGATCTGAATCTGA 462  
QY 477 TTGAGCAGGCAACCCCTGACCGGTGCGGAGAGCTGACGCGCACTTCTGCTCAATGAGC 536  
Db 463 TTGAGCAGGCAACCCCTGACCGGTGCGGAGAGCTGACGCGCACTTCTGACGAGATGAGC 522  
QY 537 GCCCGGTAGTAAGCCCCCGAGAGCCCTCTTCTTGTTCAGTTCAGAAAGGCGAGTCT 596  
Db 523 GCCGTGTAGTAAGCCCCCGAGAGCCCTTCTTGTTCAGTTCAGAAAGGAGAGAGCT 582  
QY 597 ACTTCCACCTCCATATTTCTGTGAGAGCCAGCGGGTCAAAATCCATGCTGCGCGCT 656  
Db 583 ACTTCCACATGACGCTGCTCGTGAAACCAACCGGGGTGAATCCATGCTTTTGGAGAGCTT 642  
QY 657 TCCTGAGTCAAGATTAGGGACAAAGCTGGTGAGACCATCTACCGCGGATGAGCCGAGCC 716  
Db 643 TCCTGAGTCAAGATTGCGGAAAACTGATTCAGAGAAATTACCGCGGATGAGCCGAGCTT 702  
QY 717 TGCCCAACTGTGTGCGGTGACCAAGACGCTGATGAGCGCGGAGGAGGAAACAAGTGG 776  
Db 703 TGCCAACTGTGTGCGGTGACCAAGACCAAGATGGCGCGGAGGAGGAAACAAGTGG 762  
QY 777 TGAGCAGTGTCAATCCCAACTACTCTCTGCCCCAAGACTAGCCCGAGCTGCAAGTGG 836  
Db 763 TGATGAGTGTACATCCCAATTAATTAAGCGCTGTTGAATCTCAGCGAGCTCAAGTGG 822  
QY 837 CGTGACTAATGAGAGATATATAAGCGCTGTTGAACCTGCGCGAGCGCAACGCGC 896  
Db 823 CGTGACTAATGAGAGATATATAAGCGCTGTTGAATCTCAGCGAGCTCAAGTGG 882  
QY 897 TCGTGCGCAGCAGCTGACCCACGTCAGCCAGCAGGAGAGAGAGAGAGAGATCTGA 956  
Db 883 TGGTGCGCAGCAGCTGACCCACGTCGTCAGCAGCAGGAGAGAGAGAGAGATCTGA 942  
QY 957 ACCCCAATTGTGACGCGCTGTATCCCGTCAAAAACCTCCGCGGCTATGAGAGTGG 1016  
Db 943 ATCCCAATTGTGATGCGCGGTGATCAGATCAAAAACCTTCAGCCAGTATGAGAGTGG 1002  
QY 1017 TCGGTGTGTGTGACCGGGGATCACTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1076  
Db 1003 TCGGTGTGTGTGACCGGGGATTAACCTCGAGAGAGAGAGAGAGAGAGAGAGAGAG 1062  
QY 1077 CCTGTACATCTCTTCAAGCGCGCTTCAACTCGCGGTCCAGATCAAGCGCGCTGG 1136  
Db 1063 CCTCATACATCTCTTCAATGCGCGCTTCAACTCGCGGTCCCAATCAAGAGCTGCTTG 1122  
QY 1137 ACAATGCGGAGAGATCATGCGCTGACCAAAATCCGCGCGCACTACTGTGTAAGCCCG 1196  
Db 1123 ACAATGCGGAGAGATTAATGAGCTGATTAACCGCGCGCACTACTGTGTAAGCCCG 1182  
QY 1197 CTCGCGCGCGGAGATTAAACCAACCGCATCTAACCGCATCTGAGCTGAACGGCTACG 1256  
Db 1183 AGCCCGTGAGAGCATTTCCAGCAATCGAATTAATAATTGGAATTAACCGGGTACG 1242  
QY 1257 AACCTGCTTACGCGCGCTCGCTTTCTGCGGTGGGCGGAGAGAGAGAGAGAGAGAG 1316  
Db 1243 ATCCCAATATGCGGCTTCCGCTTTCTGGAGTGGGCGGAGAGAGAGAGAGAGAGAG 1302

QY 1317 ACACCATCTGGCTGTTTGGGCGCGGCAACCAAGGAGAGAGAGAGAGAGAGAGAGAG 1376  
Db 1303 ACACCATCTGGCTGTTTGGGCGCTGCAACTACCGGAGAGAGAGAGAGAGAGAGAGAG 1362  
QY 1377 CCCAGCGCGTCCCTTCTACGCTGCTGCACTGAGCAGAGAGAGAGAGAGAGAGAGAG 1436  
Db 1363 CCCACTGTGCTCTTACGCGGTGCTAACTGAGCAGAGAGAGAGAGAGAGAGAGAGAG 1422  
QY 1437 ATTGCTGCAACAAGATGTGATCTGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1496  
Db 1423 ACTGTGCAACAAGATGTGATCTGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1482  
QY 1497 AGTCCGCAAGGCCATTCTCGCGGCAACAAGGTGCGCGTGAACCAAAAGTCAAGTGT 1556  
Db 1483 AGTCCGCAAGGCCATTCTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1542  
QY 1557 CCGCCGATGACCCCAACCCCGGTGATGCTACCTCCAAACCAACATGTGCGCGTGA 1616  
Db 1543 CCGCCGATGACCCCAACCCCGGTGATGCTACCTCCAAACCAACATGTGCGCGTGA 1602  
QY 1617 TTGACGGAGAACAGCAGCAGCCTTGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1676  
Db 1603 TTGACGGAGAACCAACGAGCCTTGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1662  
QY 1677 TTGAATCAACCCCGCTGAGAGATGACTTTGGCAAGGTGACAAAGCAGAGAGTCAAG 1736  
Db 1663 TTGAATCAACCCCGCTGAGATGACTTTGGCAAGGTGACAAAGCAGAGAGTCAAG 1722  
QY 1737 AGTCTTCCGCTGGGCGAGAGATCACTGACCGAGGTGCGGATGAGTCTACGTCAAG 1796  
Db 1723 ACTTTTCCGCTGGGCAAGAGATCACTGAGTGAAGTGAAGATTTCTACGTCAAG 1782  
QY 1797 AGGTGAGCCCAACAAAG 1856  
Db 1783 AGGTGAGCCCAAG 1842  
QY 1857 CCTGCCCTCAGTGGCGGATCCATGCACTGACGCGGAGAGAGAGAGAGAGAGAGAG 1916  
Db 1843 TGGCGGATCAGTTGGCAGCCATGACGTGACGCGGAGAGAGAGAGAGAGAGAGAG 1899  
QY 1917 CCGACAGTACCAAAACAAATGTTCTGTCAGCAGCGGAGAGAGAGAGAGAGAGAGAG 1976  
Db 1900 CAGACAGTACCAAAACAAATGTTCTGTCAGCAGCGGAGAGAGAGAGAGAGAGAGAG 1959  
QY 1977 GCAAGCATGCGAGAGATGATCAAGATTTCAACATTTGCTTCAAGCAGGAGAGAGAG 2036  
Db 1960 GCAAGCATGCGAGAGATGATCAAGATTTCAACATTTGCTTCAAGCAGGAGAGAGAG 2019  
QY 2037 ACTGTTCAAGTGTCTCCCGCGGTGTCAGATCTCAACCGGT---CGTCAGAAAGAGAG 2093  
Db 2020 ACTGTTCAAGTGTCTCC---CGTGTCAAGATCTCAACCGGT---CGTGTCAAGAGAG 2076  
QY 2094 CGTATCGAAGTCTGTGCAATTCATCTGCTGGGCGGCGCTCCGAGATGCTTGTCT 2153  
Db 2077 CGTATCGAAGTCTGTGCAATTCATCTGCTGGGCGGCGCTCCGAGATGCTTGTCT 2133  
QY 2154 CCGCTGCGATCTGTGCAACGTGAGCCTGAGTACTGTGTTCTGAGCAATAATGACTT 2213  
Db 2134 CTGCTGCGATCTGTGCAATGTGAGTTTGGATGACTGCTGCTTTGAAACAATAATGATTT 2193  
QY 2214 AAACAGGTATGCTGCGGATGTTATCTTCAAGATTTGCTGAGAGAGAGAGAGAGAG 2273  
Db 2194 AAATCAGGTATGCTGCGGATGTTATCTTCAAGATTTGCTGAGAGAGAGAGAGAGAG 2253  
QY 2274 GGCATTGCGAGTGTGGAATTGAACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2333  
Db 2254 GGAATAGAGAGTGTGAGAGCTCAAACTGGCCACAGAGAGAGAGAGAGAGAGAGAGAG 2313  
QY 2334 AAGCAG 2393  
Db 2314 CATTAAG 2373

OY	2394	GGACTCGACAAGGGGAGCCCGTCAACGCGCGGACCGGCGCCCTCGAGCAGCAAG	2453
Db	2374	GGACTCGACAAGGAGAGCCGGTCAACGAGCAGACGCGCGCCCTCGAGCAGCAAA	2433
OY	2454	GCCTACGACCAGAGCTCAAGCCGGGTGACATCCGTACTCGGTATTAACACGCCGAC	2513
Db	2434	GCCTACGACCAGAGCTCGACAGCGGAGACAAACCCGTACTCAAGTACAAACGCGCAC	2493
OY	2514	GCCGAGTTTACGAGCGCTCTGCAAGAAGATACGTCTTTGGGGGCAACCTCGGGCGACA	2573
Db	2494	GCGGAGTTTACGAGCGCTTAAAGAAGATACGTCTTTGGGGGCAACCTCGGACGACA	2553
OY	2574	GTCCTCCAGGCCAAGAAAGCGGTTCTGAACTCTCGGTCTGTTGAGGAAGCGCTAAG	2633
Db	2554	GTCCTCCAGGCCAAGAAAGAGGTTCTTGAACCTCTGGGCTGTTGAGGAACCTGTTAAG	2613
OY	2634	ACGCGCTCCTGAAAAGAAAGTCCGGTAGAGCAGTCGCCACAAGAGCCAGACTCTCTCG	2693
Db	2614	ACGCGCTCCGGAAAAAAGAGCGCCGGTAGAGCAGCTCTCTGTGAGCCAGACTCTCTCG	2673
OY	2694	GGCATCGGCAGACAGGCCAGCAGCCCGCTAAAAAGAGACTCAATTTTGGTCAGACTGGC	2753
Db	2674	GGAACCGGAAAGCGGGCCAGACAGCCTGCAAGAAAAAGATTGAATTTGGTCAGACTGGA	2733
OY	2754	GACTCAGAGTCAGTCCCGCATCCACAACCTCTCGAGAACTCCAGCAACCCCGCTGT	2813
Db	2734	GACGCAAGACTCAGTACTGACCCCCAGCCTCTCGAGACGCCACAGCAGCCCGCTGTGT	2793
OY	2814	GTGGAGCTACTACAATGGCTTCAGGCGGTGCGCACCAATGGCAGACAATAACGAAGCC	2873
Db	2794	CTGGGAATAATACGATGCTACAGGCAGTGGCGCCACAATGGCAGACAATAACGAAGCC	2853
OY	2874	GCCGACGAGTGGGTAAATGCTCAGGAAATTGGCATGGCATTTCCACATGGCTGGCGCAC	2933
Db	2854	GCCGACGAGTGGGTAAATCTCGGAAATTGGCATGGCATTTCCACATGATGGCGCAC	2913
OY	2934	AGAGTCATCACCAACAGCAGCCCGCAGCCTGGGCTTGCCCACTTCAATTAACACCTTAC	2993
Db	2914	AGAGTCATCACCAACAGCAGCCGAACTGGGCTTGCCCACTTCAATTAACACCTTAC	2973
OY	2994	AAGCAATCTCCAGTGTTCACAGGGGGCCAGCAACGACAACCACTACTTCGGCTACAGC	3053
Db	2974	AAACAATTTCCAGCCCAATCA--GAGCCTCGAACGACAATCACTACTTTGGCTACAGC	3030
OY	3054	ACCCCTGGGGTATTTGATTTCAACAGATTCCACTGSCACTTTCAACACGTAAGCTGG	3113
Db	3031	ACCCCTGGGGTATTTGACTTCAACAGATTCCACTGSCACTTTCAACACGTAAGCTGG	3090
OY	3114	CAGCGACTCATCAACAACAATGGGGATTCCGGCCCAAGAGACTCAACTTCAAACTCTTC	3173
Db	3091	CAAAAGACTCATCAACAACAATGGGGATTCCGACCCCAAGAGACTCAACTTCAAGCTCTT	3150
OY	3174	AACATCCAAGTCAAGAGGTCAAGCAGATGATGGCTCAACACCATCGTAATACTTT	3233
Db	3151	AAATATCAAGTCAAGAGGTCAAGCAGATGACGCTACGACGATTTGCCAATAACTTT	3210
OY	3234	ACCAGCAAGTTCAAGTCTTCTCGGACTCGGAGTACCACTTCCGTACGTCTCGGCTCT	3293
Db	3211	ACCAGCAAGTTCAAGTGTTTTACTGACTCGGAGTACCACTCCGTACGTCTCGGCTCG	3270
OY	3294	GCGCACCAGGGCTGCTCTCCGTTCCCGCGGAGCGTTCATGATTTCCGCAATACGGC	3353
Db	3271	GCGCATCAAGGATGCTCCGCGCTTCCAGCAGACGTCTTCATGTGCCACAGTATGGA	3330
OY	3354	TACCTGACGCTCAACAATGGCAGCCAAAGCGTGGAGCTTCATCCTTTTACTGCTGGA	3413
Db	3331	TACCTACCCCTGAACAACGGGAGTCAAGCAGTAGAGCGCTTCATTTTACTGCTGGAG	3390
OY	3414	TATTTCCCTTCTCAGATGCTGAGAAACGGGCAACAATTACCTTCAAGTACACTTTGAG	3473
Db	3391	TACTTTCTTCTCAGATGCTGCGTACCAGAAACAATTACCTTCAAGTACACTTTGAG	3450
OY	3474	GAAAGTCTTTCCACAGCAGTACGCGCACAGCCAGAGCTTGAGCCGGCTGATGATCTT	3533

Db	3451	GAAGTTCCTTCCAGCAGCTACGCTCACAGCCGAGTCTGGACCGCTCTCATGAATCCT	3510
QY	3534	CTCATCGACCAATACCTGTATTACCTGAACAGAACTCAAAATCAGTCCGGAAGTCCCA	3593
Db	3511	CTCATCGACCAATACCTGTATTACCTGAACAGAACTCAAAATCAGTCCGGAAGTCCCA	3570
QY	3594	AACAAGACTTGTGTTAGCCGTGGGTCTCCAGCTGGCATGTCTGTTAGCCCAAAAC	3653
Db	3571	CAGTCAAGGCTTCAGTTTCTCAGGCCGAGCGAGTGAATTCGGGACCACTGAGAAC	3630
QY	3654	TGGCTACCTGGACCCCTGTTATCCGGCAGCAGCCGCTTCTAAACAAAACAGACAAC	3713
Db	3631	TGGCTTCTGGACCCCTGTTATCCGGCAGCAGCCGCTTCTAAACAAAACAGACAAC	3690
QY	3714	AACAGCAATTTTACCTGGACTGGTCTTCAAAATATACTCAATGGGCGTGAATCATC	3773
Db	3691	AACAGTGAATACTCGTGACTGGAGCTACCAAGTACCACTCAATGGCAGACTCTCTG	3750
QY	3774	ATCAACCCCTGGCACTGCTATAGCCCTCACACAAGAGCAGAGCAAGTCTTTCCATG	3833
Db	3751	GTGAATCCGGCCCGGCATGGCAAGCCACAAGAGCATGAAGAAAGTTTTCCTCAG	3810
QY	3834	AGCGGTGCATGATTTTGGAAAAAGAGCGCCGAGCTTCAACACTGCATTGGACAAT	3893
Db	3811	AGCGGGGTCTCATCTTTGGGAGCAGAGGCTCAGAGAAAACAATGTGACATTGAAGA	3870
QY	3894	GTCATGATTACAGACGAGAGGAATTAAGCCACTAACCTGTGGCCACCGAAAGATT	3953
Db	3871	GTCATGATTACAGACGAGAGGAATTAAGCCACTAACCTGTGGCCACCGAAAGATT	3930
QY	3954	GGGACCGTGGCACTCAATTTCCAGAGCAGCAGACAGACCTGCGACCGAGATGTGAT	4013
Db	3931	GGTCTGTATCTACCACTCCAGAGAGGCAACAGACAAGTACCGCAGATGTCAAC	3990
QY	4014	GCTATGGAGCATTACCTGGCATGTGTGGCAAGATAGAGAGTGTACCTGAGGTCCC	4073
Db	3991	ACACAAGGCGTTCTTCAGAGCATGTGTGGCAGACAGAGATGTACTTTCAGGGGCC	4050
QY	4074	ATTGGGCCCCAAATTCCTCACACAGATGACACTTTCACCCGCTCTCTTATGGGCGG	4133
Db	4051	ATCTGGCCAAAGATTCCACACGAGCAGACATTTTCAACCTCTCCCTCATGGGTGA	4110
QY	4134	TTTGAATCAAGAACCCGCTCTCAGATCTCTCAAAAAACAGCCGTGTTCTCTGGAAT	4193
Db	4111	TTGGAATCAAGAACCCGCTCTCAGATCTCTCAAAAAACAGCCGTGTTCTCTGGAAT	4170
QY	4194	CCTCCGGCGAGTTTTCAGCTACAAAGTTTGTTCATTCTACCCCAATCTCCACAGA	4253
Db	4171	CCTTCGACCACTTTCAGTGCAGCAAGTTTGTTCATTCTACCCCAATCTCCACAGA	4230
QY	4254	CAAGTGTGTGAAAATTGAATGGAGCTGCAGAAAGAAAACAGCAAGCGCTGGAATCCC	4313
Db	4231	CAGTCAAGCGTGAATGAGTGGAGCTGCAGAAAGAAAACAGCAAGCGCTGGAATCCC	4290
QY	4314	GAAATGCAATACACATCCAATTATGCAAAATCTGCCAAGCTGATTCTTACTGTGCAAC	4373
Db	4291	GAAATGCAATACACATCCAATTATGCAAAATCTGCCAAGCTGATTCTTACTGTGCAAC	4350
QY	4374	AATGACTTTATCTAGCCCTGCCCATTTGGCACCCTTACCTTACCCTGCCCTGTAA	4433
Db	4351	AATGCGGTGTATTCAGAGCTTCCGCCCATTTGGCACCCTTACCTTACCCTGTGTAA	4410
QY	4434	TTACGTGTAAATCAATAAACCGGTTGATTGCTTCAAGTGAACCTTGGTCTCTGTCTT	4493
Db	4411	TTGCTGTAAATCAATAAACCGGTTGATTGCTTCAAGTGAACCTTGGTCTCTGTCTT	4470
QY	4494	CTTATCTTATCGTTTACCATGTATTATAGCTTACACATTAATGCTTGGTGGCTTCCG	4553
Db	4471	TCTTCTTATCTAGTTTCCATGGCTACGTAGATGAATGACATGGCGGTTAATCATTAAC	4530
QY	4554	ATAAAGACTTACGTCAATCGGGTTACCCCTAGTGAATGAGTTGCCACTCCCTCTCTGG	4613





Db 1603 TTGACGGGAACCTCAACGACCTTCCGAACACGACGCCGTTGCAAGACCGGATGTTCAAT 1662

QY 1677 TTGAACCTACCCGCCCTCTGAGCATGACTTTGGCAAGGTGAACAAGCAGGAAGTCAAG 1736

Db 1663 TTGAACCTACCCGCCCTCTGAGCATGACTTTGGCAAGGTGAACAAGCAGGAAGTCAAG 1722

QY 1737 AGTTCTTCGCTGGGCGCAGATCAAGTGAACGAGGTGGCGCATGAGTTCTACGTAGAA 1796

Db 1723 ACTTTTCCGGTGGGCAAGAGATCAAGTGTGAGGTGAGCATGATTTCTACGTCAAAA 1782

QY 1797 AGGTTGAGCCAAACAAGACCCGCCCATGACCGGATTAAGCGAAGCCCAAGCGGG 1856

Db 1783 AGGTTGAGCCAAACAAGACCCGCCCATGACCGGATTAAGCGAAGCCCAAGCGGG 1842

QY 1857 CCTGCCCTCAGTCCCGGATCCATCGACGTGACGCGGAGAGAGTCCGGTGAATTG 1916

Db 1843 TCGCGGAGTCAAGTTGCGCAGCCATCGACGTGACGCGGA -- AGCTTCGATCAACTACG 1899

QY 1917 CCGACAGGTACCAAAAACAATGTTCTGTCACGCGGGCATGCTTCAGATGCTGTTCCCT 1976

Db 1900 CAGACAGGTACCAAAAACAATGTTCTGTCACGTTGGCATGAATCTGATGCTGTTCCCT 1959

QY 1977 GCAAGACATCGAGAGAGATGAATCAAAATTTCAACATTTGCTTCAACGACGAGCAGAG 2036

Db 1960 GCAGACAAATCGAGAGAGATGAATCAAAATATCTGCTTCACTCAACGACAGAGAAAG 2019

QY 2037 ACTGTTCAAGTGTCTTCCCGCGCTGTCAGATCTCAACCCGT --- CGTCAGAAAGAGA 2093

Db 2020 ACTGTTCAAGTGTCTTCC --- CGTCAGAAATCTCAACCCGTTTCTGCTCAAAAAGG 2076

QY 2094 CGTATCGGAACTCTGTGCTATTCATCTGCTGGGCGGGCTCCCGAGATTGCTGCT 2153

Db 2077 CGTATCGGAACTGTGCTATTCATCTGCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAG 2133

QY 2154 CGGCTGCGATCTGTCAACGTGAGCTGATGATGATGATGATGATGATGATGATGATGAT 2213

Db 2134 CTGCTGCGATCTGTGCTATTCATCTGCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAG 2193

QY 2214 AAACCAAGTATGCTGCTGCGATGTTATCTTCCAGATGCTGAGAGACAACCTCTCTGAG 2273

Db 2194 AAATCAAGTATGCTGCTGCGATGTTATCTTCCAGATGCTGAGAGACAACCTCTCTGAG 2253

QY 2274 GGCATTCGAGTGTGAGTGAATGAACCTGAGCCCGGAGCCGCAAGCCCAAGCCCAAGCA 2333

Db 2254 GGAATAGACAGTGTGAGAGCTCAAACTGCGCCCAACCAACCAAGCCCGCAGAGCGG 2313

QY 2334 AAGCAGACGACGCGCGGGGTCTGCTGCTTCTGCTGCTAACAAGTACCTCGGACCTTCAAC 2393

Db 2314 CATAGGACGACGAGGAGGCTTGTGCTTCTGCTGCTAACAAGTACCTCGGACCTTCAAC 2373

QY 2394 GGCATTCGACGAGGAGGAGCCCGTCAACGCGGCGAGCGAGCGGCTCGAGCAGCAGAA 2453

Db 2374 GGCATTCGACGAGGAGGAGCCCGTCAACGAGGCGAGCGCGGCTCGAGCAGCAGAA 2433

QY 2454 GGCATTCGACGAGCTCAAAAGCGGCTGACATCCGTACCTGCGGTATTAACCAAGCCGAC 2513

Db 2434 GGCATTCGACGAGCTCAACGAGGAGCAACCCGTACCTCAAGTACCAACCAAGCCGAC 2493

QY 2514 GGCAGTTCAGAGAGCGTCTGCAAGAGATAGCTTTTGGGGCAACCTCGGCGAGCA 2573

Db 2494 GCGAGTTCAGAGAGCGCTTAAGAGATAGCTTTTGGGGCAACCTCGGCGAGCA 2553

QY 2574 GTCTTCAGGCGCAAGAGCGGCTCTGAACCTCTCGGTCTGCTGAGAGAGCGCTAG 2633

Db 2554 GTCTTCAGGCGCAAGAGCGGCTCTGAACCTCTCGGTCTGCTGAGAGAGCGCTAG 2613

QY 2634 ACGGCTCTGGAAGAAACGTCGGTGAAGCAGTCCGACACAGAGCCAGACTCTCTCTG 2693

Db 2614 ACGGCTCTGGAAGAAAGAGCGCGGTGAAGCACTCTCTGAGAGCCAGACTCTCTCTG 2673

QY 2694 GGCATTCGAGAGAGAGCGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2753

Db 2674 GGAACCGGAAAGCGGCGCAGCAGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2733

QY 2754 GACTCAGAGTCAAGTCCCGCATCCACAACCTCTCGAGAGAACTCCAGCAACCCCGCTGCT 2813

Db 2734 GACGCAAGTCAAGTCAAGTCCCGCATCCACAACCTCTCGAGAGAACTCCAGCAACCCCGCTGCT 2793

QY 2814 GTGGAGCTTACTAATGCTTCAAGCGGCTGCGCAGCAATGCGAGCAATTAAGAGAG 2873

Db 2794 CTGGAGCTAATAGATGCTTCAAGCGGCTGCGCAGCAATGCGAGCAATTAAGAGAG 2853

QY 2874 GCGGACGAGTGGGTAAATGCTTCAAGAAATTTGCAATGCGATTTCCACATGCTGGCGAC 2933

Db 2854 GCGGACGAGTGGGTAAATGCTTCAAGAAATTTGCAATGCGATTTCCACATGATGGCGAC 2913

QY 2934 AGATCATCAACCAAGCAACCCGCACTGGGCTTGCCTCACTAATAACCACTCTAC 2993

Db 2914 AGATCATCAACCAAGCAACCCGCACTGGGCTTGCCTCACTAATAACCACTCTAC 2973

QY 2994 AAGCAATCTCAGTGTCTCAACGCGGCGCAGCAACCACTACTTGGGCTACAGC 3053

Db 2974 AAGCAATCTCAGTGTCTCAACGCGGCGCAGCAACCACTACTTGGGCTACAGC 3030

QY 3054 ACCCCCTGGGATTTGATTTCAACAGATTTCACTGCGCACTTTCAACAGCTGAG 3113

Db 3031 ACCCTTGGGATTTGATTTCAACAGATTTCACTGCGCACTTTCAACAGCTGAG 3090

QY 3114 CAGCACTCATCAACCAATTTGGGATTTCCGCGCAGAGAGACTCAACTCAACTCTTC 3173

Db 3091 CAAGCACTCATCAACCAATTTGGGATTTCCGCGCAGAGAGACTCAACTCAACTCTTC 3150

QY 3174 AACATCCAAAGTCAAGAGGTCACGAGATGATGCGCTCAACAACCATCGCTAATACTT 3233

Db 3151 AACATCCAAAGTCAAGAGGTCACGAGATGATGCGCTCAACAACCATCGCTAATACTT 3210

QY 3234 ACCAGCAGGTTCAAGTCTTCTGAGTCCGAGTACCACTTCCGTACGCTGCTGCT 3293

Db 3211 ACCAGCAGGTTCAAGTCTTCTGAGTCCGAGTACCACTTCCGTACGCTGCTGCT 3270

QY 3294 GCGCAGCAGGCTGCTTCCCTCCGTTCCGCGGAGCTGTTCAATGATTTCCGCAATACGC 3353

Db 3271 GCGCAGCAGGCTGCTTCCCTCCGTTCCGCGGAGCTGTTCAATGATTTCCGCAATACGC 3330

QY 3354 TACCTAGCCTCAACATGCGAGCCAGCCGCTGAGCTTCACTTCTGCTGAGAA 3413

Db 3331 TACCTAGCCTCAACATGCGAGCCAGCCGCTGAGCTTCACTTCTGCTGAGAA 3390

QY 3414 TATTTCCCTTCTCAGATGCTGAGAGCGGCAACCTTACCTTCACTTCTGAG 3473

Db 3391 TATTTCCCTTCTCAGATGCTGAGAGCGGCAACCTTACCTTCACTTCTGAG 3450

QY 3474 GAAATGCTTCTCAGAGCTACGCGCAGAGCCAGAGCTTGAACCGGCTGATGAATCT 3533

Db 3451 GAAATGCTTCTCAGAGCTACGCGCAGAGCCAGAGCTTGAACCGGCTGATGAATCT 3510

QY 3534 CTGATCGACCAATACCTGTATTAACCTGAACAGAACTCAAAATCAGTCCGGAAGTGCCAA 3593

Db 3511 CTGATCGACCAATACCTGTATTAACCTGAACAGAACTCAAAATCAGTCCGGAAGTGCCAA 3570

QY 3594 AACAGGACTTGTGTTAGCCGTGGGTCTCAGCTGCGATGTCTGTTACGCCCAAAAC 3653

Db 3571 CAGTCAAGGCTTCAAGTCTTCAAGCCGAGCGAGTGAATTCGGAGCAGCTTAAGAAC 3630

QY 3654 TGGCTACCTGAGACCTGTTATCGGAGCAGCGGCTTCTAATAACCAAGCAACAA 3713

Db 3631 TGGCTACCTGAGACCTGTTATCGGAGCAGCGGCTTCTAATAACCAAGCAACAA 3690

QY 3714 AACAGCAATTTTACCTGAGCTGCTTCAATAATAACCTCAATGCGGCTGAATCCATC 3773

Db 3691 AACAGCAATTTTACCTGAGCTGCTTCAATAATAACCTCAATGCGGCTGAATCCATC 3750

QY 3774 ATCAACCTGCGACTGCTATGCGCTCAACAAGAGAGAGAGAGAGAGAGAGAGAGAG 3833

Db 3751 GTGAATCCGGGCGCGGCAATGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3810

QY	3834	AGCGGTGCATGATTTTGGAAAAAGAGAGCGCCCGAGCTTCAAACTGCATTGGACAAT	3893
Db	3811	AGCGGGGTTCTCATCTTTGGGAAGCAAGGCTCAGAGAAAACAAAATGTGACATTGAAAAAG	3870
QY	3894	GTCATGATTACAGACGAGGAAATTTAAAGCCACTAACCTGTGGCCACCGAAAGATT	3953
Db	3871	GTCATGATTACAGACGAGGAAATTTAAAGCCACTAACCTGTGGCCACCGAGAGATAT	3930
QY	3954	GGGACCGTGCAGTCAATTTCCAGAGCAGCAGACAGACCCTGCGACCGAGATGTGCAT	4013
Db	3931	GGTTCGTATCTACCAACTCCAGAGAGGCAACAGACAGCTACCGAGATGTCAAC	3990
QY	4014	GCTATGGAGCATTAACCTGGCATGTGTGGCAAGATAGAACGTGTACCTGCAGGGTCC	4073
Db	3991	ACACAAGGCGTCTCTCCAGGCATGTCTGGCAGGACAGAGATGTGTACTTCAAGGGGCC	4050
QY	4074	ATTGGGCCAAAATTCCCTCACACAGATGGACACTTTCACCCGCTCTCCTTATGGCGGC	4133
Db	4051	ATCTGGCAAGATTCCACACACGGAACGACATTTTCACCCCTCTCCCTCATGGGTGA	4110
QY	4134	TTTGACTCAAGAACCCGCTCTCAGATCCTCATCAAAAAACGCTGTCTCTGGCAAT	4193
Db	4111	TTGGCACTTAAACACCCCTCTCCACAGATTCTCATCAAGAACACCCCGTACTGCGAAT	4170
QY	4194	CCTCCGCGAGTTTTCAGCTACAAAGTTTGTCTTCATTCATCAACCCATATCTCCACAGA	4253
Db	4171	CCTTCGACCACCTTCAGTGCAGCAAGTTTGTCTCTTCATCACAACAGTACTCCACGGGA	4230
QY	4254	CAAGTGAGTGTGAATTTGAATGGAGCTGCAGAAAGAAAACAGCAAGCGCTGGAATCCC	4313
Db	4231	CAGTCAAGCGTGAGATCGATGGGAGCTGCAGAAAGAAAACAGCAAAACGCTGGAATCCC	4290
QY	4314	GAAATGCATACACATCCAAATTAATGCAAAATCTGCCAAGTTGATTTTACTGTGACAAC	4373
Db	4291	GAAATGCATACACTTCCAACTACACAAGTCTGTTAATGTGAACTTACTGTGACACT	4350
QY	4374	AATGCACTTATACCTGAGCCTGCCCCCATTTGGCAACCCCGTTACCTTACCCTGCCCTGTA	4433
Db	4351	AATGCGGTGATTCAGAGCCTGCCCCCATTTGGCAACCAAGATACCTGACTCGTAATCTGTA	4410
QY	4434	TTACGTGTTAATCAATAAACCCGTTGATTGCTTTCAGTTGAATTGCTCTCTGTCTT	4493
Db	4411	TTGCTGTAAATCAATAAACCCGTTAATTCGTTTTCAGTTGAATTGCTCTCTGTCTT	4470
QY	4494	CTTATCTTATCGGTTTACCATGTGTTATAGCTTACACATTAATCTGCTGTGCGCTTCCG	4553
Db	4471	TCTTCTTATCTAGTTTCCATGCTACGTAGATAAGTAGCAATGGCGGTTAATCATTAAC	4530
QY	4554	ATAAAGACTTACGTATCGGTTAACCCCTAGTAGATGAGTTGCCACTCCCTCTCTCG	4613
Db	4531	TACAAGG-----ACCCCTAGTAGATGAGTTGCCACTCCCTCTCTCG	4574
QY	4614	CGCTCGCTCGCTCGGTGGGCGCTGCGGACCAAAAGTCCGCAACGGCAGAGCTCTGTCT	4673
Db	4575	CGCTCGCTCGCTCACTGAGCGCGGCGCAACCAAAAGTCCGCGGCGCTTGTGCGG	4634
QY	4674	GCGGCGCCCAACCGAGCGAGCGCGGCAAGAGAGGAGTGGCA	4718
Db	4635	GCGGCGCTCAGTGAGCGAGCGCGGCAAGAGAGGAGTGGCA	4679

	:	NUMBER OF SEQ ID NOS:	18	
	:	SOFTWARE:	Patentin version 3.1	
	:	SEQ ID NO	12	
	:	LENGTH:	4679	
	:	TYPE:	DNA	
	:	ORGANISM:	adeno-associated virus 2	
	:	US-10-038-	972A-12	

  

Query Match	Best Local Similarity	Score	DB	Length
Matches 3790; Conservative	80.2%; Pred. No. 0;	3107.8;	15;	4679;
	Mismatches	882;	Indels	53;
	Gaps			10

  

QY	1	TTGCCACTCCCTCTGCGCGCTCGTCTCGTGCGGTGGCCACCAAGTCCGC	60
	1	TTGGCCACTCCTCTGCGCGCTCGTCTCGTCACTGAGCCCGCGCACCAAGTCGCC	60
Db			
QY	61	AGACGGCAGAGCTGTGCTGTGCCGCCGCCACCAGCAGCGAGCGCGCAGAGGGAGTG	120
Db	61	CGACGCCCCGGCTTTGCCCGCGCGCTCAGTAGCGAGCGCGCAGAGAGGAGTG	120
QY	121	GGAACAATCATCTAGGGGTAATCGCGAAGCGCTCCCAAGCTGCGCGTCAAGCGCTGA	180
Db	121	GCCAATCATCTAGGGGT-----TCTTGAGGGGTGAGTCTGTGA	163
QY	181	CGTAAATTACGTATAGG---GAGTGTCTGTATTAGCTGTCACTGAGTCTTTGC	237
Db	164	CGTGAATTACGTATAGGTTAGGAGGTCTGTATTAGAGTCACTGAGTG-TTTGC	222
QY	238	GACATTTTGCAGACACCACTGCGCATTTAGGGTATATATGCGCAGTGAGCGAGCAGAT	297
Db	223	GACATTTTGCAGACACCATGTGTCACTGTGGTATTAAAGCCGAGTGAGCAGCAGGT	282
QY	298	CTCCATTTTG-AACGCGAAATTGAACGAGCAGCAGCCATGCCGGCTTCTACGAGATCG	356
Db	283	CTCCATTTTGAAGCGGAGGTTGAACGCGCAGCGCCATGCCGGGTTTTACGAGATTG	342
QY	357	TGATCAAGGTGCCGAGCGCACTGACGAGCAGCAGTCCCGGCAATTCTGACTCGTTGTGA	416
Db	343	TGATTAAGTCCCAGCGACTTTGACGAGCATCTGCCCGCATTTCTGACAGCTTTGTGA	402
QY	417	GCTGGGTGGCCGAGAAGGAATGGAGCTGCCCCCGGATTTCTGACATGGAATCTGA	476
Db	403	ACTGGGTGGCCGAGAAGGAATGGAGTTGCCCGCAGATTCTGACATGGAATCTGA	462
QY	477	TTGAGCAGGCAACCCCTGACCGTGGCCGAGAAGCTGACGCGCATCTCTGTCTCAATGCG	536
Db	463	TTGAGCAGGCAACCCCTGACCGTGGCCGAGAAGCTGACGCGCATCTCTGACGGAATGCG	522
QY	537	GCCGCGTAGTAAGGCCCGGAGGCCCTCTTTCTTTGTTCACTTCAGTTCAGAAAGGCGAGTCT	596
Db	523	GCCGCGTAGTAAGGCCCGGAGGCCCTTTCTTTGTTCACTTCAGTTCAGAAAGGAGAGCT	582
QY	597	ACTTCCACCTCATATTCTGTGGAGAACACGCGGGTCAAATCCATGTGCTGGCGCT	656
Db	583	ACTTCCACATGACGCTGTCTGTGAACACACCGGGTGAATCCATGTGTTTGGGACGTT	642
QY	657	TCCTGAGTCAGATTAGGGAACAAGCTGTGTCAGACCATCTACCGCGGATCGAGCCGACCC	716
Db	643	TCCTGAGTCAGATTCCGGAATACTGATTCAAGAAATTTACCGCGGATCGAGCCGACTT	702
QY	717	TGCCCAACTGTTGCGCGGTGACCAAGACGCGTAATGGCGCGGAGGGGGAACAAGGTGG	776
Db	703	TGCCAAACTGTTGCGCGGTGACCAAGACCAAGAAATGGCGCGGAGGGGGAACAAGGTGG	762
QY	777	TGACAGAGTGCTACATCCCACTAATCTCTGCCCCAAGACTCAGCCGAGCTGCAAGTGG	836
Db	763	TGATGAGTGCTACATCCCACTAATCTGCTCCCAAAACCAAGCTGAGCTCCAGTGG	822
QY	837	CGTGAATAATGAGAGGATATAAGCGCTGTTGAACCTGCGCGAGCGCAAAACGCG	896
Db	823	CGTGAATAATGAGACAGATTTAAAGCGCTGTTGAATCTCAAGGAGCGTAAACGCT	882
QY	897	TCGTGGCGCAGACCTGACCCAGCTCAGCCAGAGCCAGAGCAGAACAGAGAAATCTGA	956



Db 883 TGGTGGCGCAGCATCTGACGCACGTCGCGCAGACGAGAGCAACAAGATCAGA 942

QY 957 ACCCAATTCTGACGCGCTGTCTATCCCGTCAAAAACCTCCGCGCGCTAATGAGCTGG 1016

Db 943 ATCCCAATTCTGATGCGCCGGTGATCAGATCAAAAACCTCAGCCAGTACATGAGCTGG 1002

QY 1017 TCGGGTGGCTGTGGAACGGGGCATCCTCCGAGAGCAGTGGATCCAGAGGAGCCAGG 1076

Db 1003 TCGGGTGGCTGTGGAACAGGGGATTACTTCGAGAGACAGTGGATCCAGAGGAGCCAGG 1062

QY 1077 CCTCGTACATCTCCTTCAACGCGCTTCAACTCGCGGTCCAGATCAAGCCGCTGG 1136

Db 1063 CCTCATACATCTCCTTCAATGCGGCTCAACTCGCGGTCCCAATCAAGGCTGCTGG 1122

QY 1137 ACAATGCCGGAAGATCATGGCGCTGACCAAAATCCGCGCCGACTACCTGTAGCCCCG 1196

Db 1123 ACAATGCCGGAAGATTATGAGCCTGACTAAACCGCCCCGACTACCTGTGGGCCAGC 1182

QY 1197 CTCCGCGCGCGCATTTAAACCAACCGCATCTACCGCATCCTGAGCTGAACGGCTACG 1256

Db 1183 AGCCCGTGGAGCATTTCCAGCAATCGGATTTATAAATTTGGAATAAACGGGTACG 1242

QY 1257 AACCTGCTTACGCGCGCTCCGCTTTCTCGGCTGGGCCAGAAAAGTTGGGAAGCGCA 1316

Db 1243 ATCCCAATATGCGGCTTCCGCTTTGAGATGGGCCACGAAAAGTTGGCAAGAGA 1302

QY 1317 ACACCATCTGCTGTTGGGCCGCGCACCAAGGCAAGACCAACATCGCGAGGCCATCG 1376

Db 1303 ACACCATCTGCTGTTGGGCCCTGCAACTACCGGGAAGACCAACATCGCGAGGCCATAG 1362

QY 1377 CCACGCGCTGCTTCTTCTACGCGCTGCGCACTGGAACCAATGAGAATTTCCCTCAATG 1436

Db 1363 CCACCATCTGCTTCTTCTACGCGGTGCGTAACTGGAACCAATGAGAATTTCCCTCAACG 1422

QY 1437 ATTGCTGCAAGATGCTGATCTGTGGAGAGGGAAGATGACGCGCAAGTCTGG 1496

Db 1423 ACTGTGTCGAAGAAGATGCTGATCTGTGGAGAGGGAAGATGACGCGCAAGTCTGG 1482

QY 1497 AGTCCGCCAAGGCCATTTCTCGGCGGAGCAAGGTGCGGTCGACCAAAAGTCAAGTCTG 1556

Db 1483 AGTCCGCCAAGGCCATTTCTCGGAGGAAGCAAGGTGCGGTCGACCAAAAGTCAAGTCTG 1542

QY 1557 CCGGCCAGATGACCCCAACCCCGTGTGCTCACTCCCAACCAACATGTGCGCGCTGA 1616

Db 1543 CCGGCCAGATGACCCCACTCCCGTGTGCTCACTCCCAACCAACATGTGCGCGCTGA 1602

QY 1617 TTGACGGGAACAGCAACCACTTGCAGACAGACGCGTTGCAGAGCCGATGTTCAAT 1676

Db 1603 TTGACGGGAACCAACGACCTTGCAGACAGACGCGTTGCAGAGCCGATGTTCAAT 1662

QY 1677 TTGAACCTACCCGCGCTGTGAGCATGCTTTGGCAAGGTGACAAAGCAAGAAGTCAAAG 1736

Db 1663 TTGAACCTACCCGCGCTGTGATCATGACTTTGGGAAGTCAACCAAGAAGTCAAAG 1722

QY 1737 AGTTCCTCCGCTGGGCGCAGGATCAGCTGACCGAGGTGGCGCATGATTCTACGTCAGAA 1796

Db 1723 ACTTTTCCGCTGGGCAAAAGATCAGTGTGAGGTGAGCATGAATTTCTACGTCAGAAA 1782

QY 1797 AGGTGAGGCAACAAAAGACCCCGCATGACCGGATGAACCGATTAAGCGAACGCGG 1856

Db 1783 AGGTGAGGCAACAAAAGACCCCGCATGACCGGATGAACCGATTAAGCGAACGCGG 1842

QY 1857 CCTGCCCTCAGTCCGCGATCCATCAGCTCAGACGCGGAAGAGCTCCGTTGACTTTG 1916

Db 1843 TGGCGAGTCACTGCGCAGCCATCAGCTCAGACGCGGA--AGCTTCGATCAACTACG 1899

QY 1917 CCGACAGTACCAAAAGAAATGTTCTGCTACGCGGCGATGCTTCAATGCTGTTCCCT 1976

Db 1900 CAGACAGTACCAAAAGAAATGTTCTGCTACGCTGGCGATGAATGCTGTTCCCT 1959

QY 1977 GCAAGACATGCGAGAGATGAATCAGAAATTTCAACATTTGCTTCAAGCAAGGAGCAGAG 2036

Db 1960 GCAGCAATGCGAGAGATGAATCAGAAATTCAAATATCTGCTTCACTCAGCAGACGAAG 2019

QY 2037 ACTGTTCAAGTGTCTTCCCGCGGTGTCAAGATCTCAACCGGT---CGTCAGAAAGAGA 2093

Db 2020 ACTGTTAGAGTGTCTTCC---CGTGTCAAGATCTCAACCGGTTCTGCTGTCAAAAGG 2076

QY 2094 CGTATCGGAACCTCTGCGCATTCATCATCTGCTGGGCGGCTCCGAGATGCTTGTCT 2153

Db 2077 CGTATCAGAACTGTGCTACATTCATCATATCATGGAAGAGTGCCAGA---CGTTGCA 2133

QY 2154 CCGCTTCCGATCTGTGTCAACGTGGAACCTGATGACTGTGTTCTGACCAATAATGACTT 2213

Db 2134 CTGCTTCCGATCTGTGTCAATGTGAATTTGATGACTGTGATCTTGAACAATAATGATTT 2193

QY 2214 AAACCAAGTATGCTGCGCATGCTTATCTTCCAGATTTGCTGAGACAACTCTCTAG 2273

Db 2194 AAATCAGTATGCTGCGCATGCTTATCTTCCAGATTTGCTGAGACACTCTCTGAA 2253

QY 2274 GGCATTCGCGAGTGTGGACTTGAACCTGAGCCCGCAAGCCCAAGCCACAGCA 2333

Db 2254 GGAATTAAGACAGTGTGAGAGCTCAAACTGCGCCACCAACCAAGCCGAGCGG 2313

QY 2334 AAGCAAGACGACGCGCGGTCTGTGCTTCTGCTTCAAGTACCTCGACCTTCAAC 2393

Db 2314 CATTAAGACGACGAGGCTTGTGCTTCTGCTTCAAGTACCTCGACCTTCAAC 2373

QY 2394 GGAATTCGACGAGGCGGCGCTTCAACGCGCGGACGCGCGCTCGAGCAGCAGAA 2453

Db 2374 GGAATTCGACGAGGCGGCGCTTCAACGCGCGGACGCGCGCTCGAGCAGCAGAA 2433

QY 2454 GCTTACGACCGAGCTCAAAAGCGGTGACCAATCCGTACTGCGTATTAACACGCGAC 2513

Db 2434 GCTTACGACCGGAGCTCGACGCGGAGACCAACCGTACTCAAGTACCAACGCGAC 2493

QY 2514 GCCGATTTCAAGGAGCTCTGCAAGAAAGATACGCTTTTGGGGCAACCTCGGCGAGCA 2573

Db 2494 GCGGATTTCAAGGAGCGCTTAAAGAAAGATACGCTTTTGGGGCAACCTCGGCGAGCA 2553

QY 2574 GTCTTCCAGGCGCAAGAGCGGTTCTCGAACCTCTCGTCTGTGAGGAAGCGCTAAG 2633

Db 2554 GTCTTCCAGGCGCAAGAGGTTCTTGAACCTCTGGGCTGTGAGGAACCTGTTAAG 2613

QY 2634 ACGGCTCTGGAAGAAAGCGTCCGTAGAGCAGTCCGCCAAGAGCCAGACTCTCTCG 2693

Db 2614 ACGGCTCTGGAAGAAAGCGGCTAGAGCAGTCTCTGTGAGCCAGACTCTCTCG 2673

QY 2694 GGCATCGGCAAGACGAGCGCAGACCCGCTAAAAAGACTCAATTTTGTGACTGGC 2753

Db 2674 GGAACCGAAAGCGGCGCAGACCTGCAAGAAAGATGAATTTTGTGACTGGC 2733

QY 2754 GACTCAGATCAGTCCCGGATTCACAACCTCTCGAGGAACCTCAGCAACCCCGCTGT 2813

Db 2734 GAGCAGACTCAGTACCTGACCCCGAGCTCTCGGACAGCCACGAGCCCGCTGTGT 2793

QY 2814 GTGGACCTTACTACAATGCTTCAAGCGGTGGCGCAACAATGGCAGACATAAGAGCC 2873

Db 2794 CTGGGAATAATAAGATGCTTCAAGCGAGTGGCGCAACAATGGCAGACATAAGAGCC 2853

QY 2874 GCCGACGAGTGGTATGCTTCAAGGAATTTGGCATTTCCACATGCTGGCGGAC 2933

Db 2854 GCCGACGAGTGGTATGCTTCTCGGAATTTGGCATTTCCACATGATGGCGGAC 2913

QY 2934 AGAGTATCAACCAAGACCCGCAACCTGGGCTTTGCCCACTTCAATTAACACCTCTAC 2993

Db 2914 AGAGTATCAACCAAGACCCGCAACCTGGGCTTTGCCCACTTCAATTAACACCTCTAC 2973

QY 2994 AAGCAATCTCAGTGTCTTCAACGCGGCGCAGCAACCACTACTTGGCTACAGC 3053

Db 2974 AAACAAATTTCCAGCCAACTCA---GGAGCTCGAAGCACAATCACTTGGCTACAGC 3030

QY 3054 ACCCCCTGGGGTATTTGATTTCAACAGATTTCACTGCACTTTTCAACGAGTCTGG 3113

Db 3031 ACCCCTGGGGTATTTGATTTCAACAGATTTCACTGCACTTTTCAACGAGTCTGG 3090



OY	3114	CAGGCACTCATCAACAACAATTTGGGGATTCCGGCCCAAGAGACTCAACTCAAACTCTTC	3173
Db	3091	CAAGAAGCTCATCAACAACAACACTGGGGGATTTCCGACCCCAAGAGACTCAACTCAAGCTCTTT	3150
OY	3174	AACATCCAAAGTCAAGAGAGTCACGACGAATGATGGCGCTCAACAACATCGTAATAACCTT	3233
Db	3151	AACATTCAAGTCAAGAGAGTCACGAGAATGACGGTAGCAGACGATGGCCAATAACCTT	3210
OY	3234	ACCAGCAGCGTTCAAGTCTTCTCGGACTCGAGTACCAGCTTCCGTACGTCCTCGGCTCT	3293
Db	3211	ACCAGCAGCGTTCAAGTGTTTACTGACTCGAGTACCAGCTTCCGTAACGTCCTCGGCTCG	3270
OY	3294	GCGCACCAGGCGTCCCTCCGTTCCCGCGGAGCGTGTTCATGATTTCCGCAATACGGC	3353
Db	3271	GCGCATCAAGGATGCTCCCGCGGTTCCAGCAGACGCTTTCATGGTGCCACAGTATGGA	3330
OY	3354	TACCTGACGCTCAACAATGGCAGCCAGCCGTGGACGTTCACTCTTTTACTGCTGGAA	3413
Db	3331	TACCTCACCCCTGAACAACGGGAGTCAAGGAGTAGGACGCTCTCATTTTACTGCTGGAG	3390
OY	3414	TATTTCCCTTCTCAGATGCTGGAACGGGGCAACAATTACTTTCAGCTACACCTTTGAG	3473
Db	3391	TACTTTCCTTCTCAGATGCTGCGTACCAGAAACAATTACTTTCAGCTACACTTTTGAG	3450
OY	3474	GAAGTGCTTTCCACAGCAGCTACGCGCACAGCCAGAGCTTGACCGGCTGATGAATCCT	3533
Db	3451	GACGTTCCCTTCCACAGCAGCTACGCTCACAGCCAGAGCTTGACCGCTCATGAATCCT	3510
OY	3534	CTCATGCAACCAATACCTGTATTAACCTGAACAGAACTCAAAATCAGTCCGGAAGTGCCTAA	3593
Db	3511	CTCATGCAACCAATACCTGTATTAACCTGAACAGAACTCAAAATCAGTCCGGAACCAACAG	3570
OY	3594	AACAAGGACTGTGTTTAGCCGCGGTCTCCAGCTGGCATGTCTGTTACGCCCAAAAC	3653
Db	3571	CAGTCAAGGCTTCAGTTTCTCAGGCGGAGCGAGTGACATTCGGGACCAGCTAGGAAC	3630
OY	3654	TGGCTACCTGGACCCGTGTTATCGGCAGCAGCGCGTTTCTAAACAACAAACAGACAACAAC	3713
Db	3631	TGGCTTCTGGACCCGTGTTATCGGCAGCAGCGAGTATCAAAACATCTGCGGATAACAAC	3690
OY	3714	AACAGCAATTTTACCTGGAAGTGTGCTTCAAATATTAACCTGAATGGCGCTGAATCCATC	3773
Db	3691	AACAGTGAATTAATCTGTGGAAGTGAAGTACCAAGTACCACTCAATGGCAGAGACTCTCTG	3750
OY	3774	ATCAACCCCTGGCACTGCTATGGCCTCAACAAGAAGCAGCAAGTCTTCTCCCATG	3833
Db	3751	GTTGAATCCGGGCCCGGCATGGCAAGCCACAAGGACGATGAAGAAAGTTTTCCTCAG	3810
OY	3834	AGCGGTGCATGATTTTGGAAAAAGAGAGCGCCGGAGCTTCAAACACTGCATTGGACAAT	3893
Db	3811	AGCGGGGTTCATCTTTGGGAAGCAAGGCTCAGAGAAAAAATAATGTGACATTGAAAAAG	3870
OY	3894	GTCATGATTAACAGACGAAGAGAAATTAAGCCACTAACCCCTGTGGCCACCGAAAGATT	3953
Db	3871	GTCATGATTAACAGACGAAGAGAAATCAAGACAACCAATCCCTGTGGCTACGAGCAGTAT	3930
OY	3954	GGGACCGTGGCAGTCAATTTCCAGACGACGACAGACAGACCCTGCGACCGAGATGTGCAT	4013
Db	3931	GGTTCGTATTAACCAACCTCCAGAGAGGCAACAGACAAGCAGTACCGCAGATGTCAAC	3990
OY	4014	GCTATGGAGCATTAACCTGGCATGTGTGGCAAGATAGAGACGTGTACTTCGACGGTCCC	4073
Db	3991	ACACAAGGCGTTCTTCCAGGCGATGTCTGGCAGGACAGAGATGTACTTTCAGGGGCC	4050
OY	4074	ATTGGGCCCCAAATTCCTCAACAGATGGAACAATTCAACCCGTCTCTTATTTGGGGCGC	4133
Db	4051	ATCTGGGCAAAAGATTCCAACACGGGACGGACATTTTCAACCCCTCTCCCCCTCATGGGTGA	4110
OY	4134	TTTGAAGTCAAGAACCCGCTCTCAGATCTCATCAAAAAACAGCCTGTTCTGCGAAT	4193
Db	4111	TTGGAAGTAAACACCCCTCTCCACAGATTTCTATCAAGAACACCCGCTACCTGCGAAT	4170

QY		4194	CCTCCGGCGAGTTTTCAGCTACAAAGTTTGCTTCATTCATCACCCAATTACTCCACAGA	4253
Db		4171	CCTTCGACCACCTTCAGTGC GGCAAGTTTGCTTCATTCACAGTA CTCCACGGGA	4230
QY		4254	CAAGTAGTGTGAAATTGNAATGGGAGCTGCAGAAAAGAAAA CAGCAAGCGCTGGAATCCC	4313
Db		4231	CAGGTCA GGTGAGATCGAGTGGGAGCTGCAGAAAGAAAA CAGCAAACGCTGGAAATCCC	4290
QY		4314	GAA GTGAGTACACATCCAA TTATGCAAAATCTGCCAAGTTTACTGTGACAAC	4373
Db		4291	GAAATTCA GTACACTTCCA ACTACACA CAAGTCTGTTAATGTGACTTACTGTGACACT	4350
QY		4374	AATGAC TTTATACTGAGCCTCGCCCCATTGGCACCCGTTACCTTACCCGCTCCCTGTAA	4433
Db		4351	AATGCGGTGATTCAGAGCCTCGCCCCATTGGCACCA GATACTGACTCGTAATCTGTAA	4410
QY		4434	TTACGTGTTAATCAATAAACCGTTGATTCGTTTCAGTTGAAC TTGGTCTCTGTCCTT	4493
Db		4411	TTCGTTGTTAATCAATAAACCGTTTAATCGTTTCAGTTGAAC TTGGTCTCTGCGTATT	4470
QY		4494	CTTATCTTATCGGTTACCATG GTTATAGCTTACACATTA CTGCTTGTTGCGTTGCG	4553
Db		4471	TCTTCTTATCTAGTTCCATGGCTAGTAGATAAGTACATGGCGGTTAATCATTAAC	4530
QY		4554	ATAAAGA CTTACGTCATCGGGTTACCCCTAGTGA TGAGTTGCCACTCCCTCTGCG	4613
Db		4531	TACAAGGA-----ACCCTAGTGA TGAGTTGGCCACTCCCTCTGCG	4574
QY		4614	CGCTCGCTCGCTCGGTGGGCGCTGCGGACCA AAGTCCGACAGACGGCAGAGCTTGCTCT	4673
Db		4575	CGCTCGCTCGCTCACTAGAGCGCGGCGACCA AAGTCCGCCAGCGCGGCTTTGCGCG	4634
QY		4674	GCTGGCCCCACCGAGCGAGCGGCGCAGAGAGGGAGTGGGCCAA	4718
Db		4635	GCGCGGCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGGCCAA	4679

RESULT 4  
US-10-077-294-1

Sequence 1, Application US/10077294

Publication No. US20020159979A1

GENERAL INFORMATION:

APPLICANT: Johnson, Phillip R.

TITLE OF INVENTION: Adeno-Associated Virus Materials and

## Methods

NUMBER OF SEQUENCES: 3

CORRESPONDENCE ADDRESS:

ADDRESSEE: Marshall, O'Toole, Gerstein, Murra  
 STREET: 6300 80th Tower 323 S Wacker Drive

STREET: 6300  
CITY: Chicago

CITY: Chicago  
STATE: Illinois

STATE: ILLINOIS  
COUNTRY: USA

COUNTRY: USA  
ZTP: 60606

COMPUTER READABLE FORM:  
ZIP: 80808

```

CONFIRM RELOADABLE CONF:
MEDIUM TYPE: Floppy disk

```

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/077,294

FILING DATE: 15-Feb-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/691,604

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: No. US20020159979A1

REGISTRATION NUMBER: 35,302  
 PENDING/DOCKET NUMBER: 31075

REFERENCE/DOCKET NUMBER: 31975  
TELECOMMUNICATION INFORMATION:

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (313) 474-6300

TELEPHONE: (312) 474-63  
TELEFAX: (312) 474-0448

TELEFAX: (312)  
TELEFAX: 25-3856

95895-27 : YETTEL

INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4680 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
us-10-077-294-1

Query Match 65.3%; Score 3082.4; DB 15; Length 4680;  
Best Local Similarity 80.0%; Pred. No. 0;  
Matches 3781; Conservative 0; Mismatches 891; Indels 54; Gaps 11;

QY 1 TTGCCCCACTCTCTCTGCGCGCTCGCTGCGGTGGGCGCTCGGACCAAGTCCGC 60  
DB 1 TTGGCCACTCCTCTCTGCGCGCTCGCTGCTCACTGAGCGCGGCGACCAAGTCCGC 60  
QY 61 AGACGGCAGAGCTCTGCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120  
DB 61 CGACGCCCGCGCTTTGCCCGCGCGCTGAGCGAGCGCGCGCGCGCGAGAGGAGTG 120  
QY 121 GGCACCTCCATCACTAGGGTAATCGCGAAGCGCTCCACGCTGCCGCTGACGCTGA 180  
DB 121 GCCAATCCATCACTAGGGT-----TCCGAGAGGGGTGAGTCCGTA 163  
QY 181 CGTAATACGTATAGGG--GAGTGTCTGTATTAAGTGTCACTGAGTGTCTTTGC 237  
DB 164 CGTGAATACGTATAGGGTAGGGAGGTCTGTATTAGAGGTCACTGAGTG-TTTTGC 222  
QY 238 GACATTTTGGGACACCAAGTGGCCATTAGGGTATATATGCGCGAGTGAGCGAGAT 297  
DB 223 GACATTTTGGGACACCATGTGTGTCACGCTGGGATTAAAGCCGAGTGAAGCAGCAGGT 282  
QY 298 CTCCATTTTG-ACCGGAAATTGAACGAGCAGCAGCCATGCGCGCTTCTACGAGATCG 356  
DB 283 CTCCATTTTGAAGCGGAGTTTGAACGCGCAGCCGCGCATGCGGGGTTTACGAGATTG 342  
QY 357 TGATCAAGTGGCCGAGCGACCTGAGCAGCACTGCGGCGCATTTCTGACTCGTTTGTGA 416  
DB 343 TGATTAAGTCCCCAGCGACCTTGAAGGGGATCTGCCGCGCATTTCTGACAGCTTTGTGA 402  
QY 417 GCTGGGTGGCCGAGAGGATGGAGCTGCCCCCGGATTTCTGACATGGATCTGAATCTGA 476  
DB 403 ACTGGGTGGCCGAGAGGATGGAGTTGCGCGCAGATTCTGACATGGATCTGAATCTGA 462  
QY 477 TTGAGCAGGACCCCTGACCGGTGGCGGAGAGCTGACGCGCGACTTCTGTCCAATGGC 536  
DB 463 TTGAGCAGGACCCCTGACCGGTGGCGGAGAGCTGACGCGCGACTTCTGACGGAATGGC 522  
QY 537 GCGCGGTGAAGAGCCCGGAGCGCTCTTCTTGTTCAGTTGAGAAAGGCGAGTCTCT 596  
DB 523 GCGGTGAAGAGCCCGGAGCGCTTCTTGTTCAGTTTGAAGAGGAGAGAGCT 582  
QY 597 ACTTCCACTCCATATTCTGTGAGACCAAGGGGTCAAAATCCATGTGTGCGCGCT 656  
DB 583 ACTTCCACTGACAGTCTGTGTAACCAAGGGGTGAATCCATGTGTGCGAGCTT 642  
QY 657 TCCTGAGTCAATTAGGACAGCTGTGAGACCACTTACCGCGGATCGAGCGGACCC 716  
DB 643 TCCTGAGTCAATTTCGCGAATACTGATTCAGAGATTACCGCGGATCGAGCGGACTT 702  
QY 717 TGCCCACTGCTTGGCGGTGACCAAGCGGTATGGCGCGGAGGGGGAACAAGGTGG 776  
DB 703 TGCCCACTGCTTGGCGGTGACCAAGCGGTATGGCGCGGAGGGGGAACAAGGTGG 762  
QY 777 TGAGCAGTGTACATCCCAACTACTCTGCGCAAGACTCAGCCGAGCTGAGTGGG 836  
DB 763 TGAGTGTGTACATCCCAATTACTTGTCTCCCAAAACCCAGCTGAGCTCCAGTGGG 822  
QY 837 CGTGAATAATGAGAGATATATAAGCGCTGTTGAACCTGCGCGAGCGCAAAAGCGC 896  
DB 823 CGTGAATAATGAGAGATATATAAGCGCTGTTGAACCTGCGCGAGCGCAAAAGCGT 882

QY 897 TCGTGGCGCAGCACCCTGACCCAGCTCAGCCAGACCCAGAGCAGAAACAAGAGAAATCTGA 956  
DB 883 TGTGGCGCAGCATCTGACCGCAGTGTGCGAGACCGCAGAGCAGAAACAAGAGAAATCAGA 942  
QY 957 ACCCAATTCTGACGCGCTGTATCCGGTCAAAAAACCTCCGCGCTACATGAGCTGG 1016  
DB 943 ATCCCAATTCTATGCGCGCTGTATCAGATCAAAAACTTCAGCCAGGTACATGAGCTGG 1002  
QY 1017 TCGGTGGCTGTGAGCCGGGATCACTCCGAGAGAGTGGATCCAGAGGAGCAGG 1076  
DB 1003 TCGGTGGCTGTGAGCAAGGGGATTACCTCGAGAGAGTGGATCCAGAGGAGCAGG 1062  
QY 1077 CCTGTACATCTCTTCAAGCGCTTCCAACTCGCGGTCCAGATCAAGCGCTCTGG 1136  
DB 1063 CCTGTACATCTCTTCAATGGCGCTCCAACTCGCGGTCCCAATCAAGCTGCTGG 1122  
QY 1137 ACAATGCCGGAAGATCATGGGCTGACCAATCCGCGCCGACTACTGTAGGCCCG 1196  
DB 1123 ACAATGCCGGAAGATTATGAGCTGACTAAACCGCGCCGACTACTGTAGGCCAGC 1182  
QY 1197 CTGCGCGCGCGGACATTAAACCAACCGCATCTACCGCATCTGAGCTGAACGGCTACG 1256  
DB 1183 AGCCGTGAGAGACATTTCAGCAATCGATTATTAATAATTGGAATCAACGGGTACG 1242  
QY 1257 AACCTGCTACGCGCGCTCCGTCTTCTCGGCTGGCGCCAGAAAGGTTCCGGAAGCGCA 1316  
DB 1243 ATCCCAATATGCGCTTCCGTCTTCTGAGTGGCGCCAGAAAGTTCCGCAAGAGGA 1302  
QY 1317 ACACATCTGGCTGTTTGGCGCGCCACCAAGCGGCAACCAATCGCGGAAGCCATCG 1376  
DB 1303 ACACATCTGGCTGTTTGGCGCTGCACTACCGGGAAGCAACATCGCGGAGCCATAG 1362  
QY 1377 CCGAGCGCGTGCCTTCTACGCGCTGCGTCACTGAGCAATGAGAACTTCCCTTCAATG 1436  
DB 1363 CCGACACTGTGCCCTTCTACGCGGTGCGTAACTGAGCAATGAGAACTTCCCTTCAACG 1422  
QY 1437 ATTGCTGCAACAAGATGTGTATCTGTGGAGAGAGGCAAGTACCGGCAAGTCTGG 1496  
DB 1423 ACTGTGTCAACAAGATGTGTATCTGTGGAGAGAGGGAAGTACCGGCAAGTCTGG 1482  
QY 1497 AGTCGCCAAGGCCATTCTCGCGCGCAGCAAGGTGCGGTGACCAAAAGTCAAGTCT 1556  
DB 1483 AGTCGCCAAGGCCATTCTCGAGAGAGCAAGGTGCGGTGACCAAAATGCAAGTCT 1542  
QY 1557 CCGCCAGATGACCCCAAGCGGTGATGTCACCTCCAACCAACATGTGCGCGTGA 1616  
DB 1543 CCGCCAGATGACCCCAAGCGGTGATGTCACCTCCAACCAACATGTGCGCGTGA 1602  
QY 1617 TTGACGGGAACAGCACCACTTTCAGACACCAAGAGCGGTTGACAGACCGGATGTTCAAT 1676  
DB 1603 TTGACGGGAACCTCAACGACTTTCAGACACCAAGAGCGGTTGACAGACCGGATGTTCAAT 1662  
QY 1677 TTGAATCAACCGCGCTGTGAGCATGACTTTGGCAAGGTGACAAAGCAGGAAGTCAAG 1736  
DB 1663 TTGAATCAACCGCGCTGTGATCATGACTTTGGGAAGGTGACCAAGCAGGAAGTCAAG 1722  
QY 1737 AGTTCTTCGCTGGCGCAGATCACTGACCGAGGTGCGCATGAGTTCTACGTCAAG 1796  
DB 1723 ACTTTTCCGCTGGCAAGGATCACTGTTGAGGTGAGCATGAATTTACGTCAAA 1782  
QY 1797 AGGTGGAGCCAAACAAAGACCCGCGCGATGACGCGGATAAAGCGAGCCCAAGCGGG 1856  
DB 1783 AGGTGGAGCCAAAGAAAGACCCGCGCGATGACGCGGATATAAGTGAAGCCCAAGCGGG 1842  
QY 1857 CTTGCCCTCAGTGGCGGATCCATGAGCTCAGACGCGGAGAGAGCTCCGCTGACTTG 1916  
DB 1843 TGGCGAGTCAAGTGGCGGATCCATGAGCTCAGACGCGGAG--AGCTTCATCACTACG 1899  
QY 1917 CCGACAGGTACCAAAACAATGTTCTGTCACGCGGCGATGCTTCAGATGCTGTTCCCT 1976  
DB 1900 CAGACAGGTACCAAAACAATGTTCTGTCACGCGGCGATGAATCTGATGCTGTTCCCT 1959





Db	4111	TTGGCACTTAACACACCCTCCTCCACAGATTCTCATCAAGAACACCCCGGTACTCGCAAT	41170
OY	4194	CCTCCGGCGGAGTTTTCAGCTACAAAGTTTGCTTCATTGATCACCCAATACTCCACAGGA	4253
Db	4171	CCTTCGACCACCTTCAGTGCAGCAAGTTTGCTTCCTTCATCAACAGTACTCCACGGGA	4230
OY	4254	CA-AGTAGTGTGGAAATTGAATGGGAGCTGCAGAAAGAAAACAGCAAGCGCTGGAATCC	4312
Db	4231	CACGGTCAGCGTGAGATCGAGTGGGAGCTGCAGAAAGAAAACAGCAACGCTGGAATCC	4290
OY	4313	CGAAGTCAGTACACATCCAAATTATGCAAAATCTGCCAACGTTGATTTTACTGTGGACA	4372
Db	4291	CGAAATTACGTACACTTCCAACTACACAACAAGTCTGTTAATCGTGAAGTTAACCGTGATAC	4350
OY	4373	CAATGACTTTATACTGAGCCTCGCCCCCATTTGGCAACCCGTTAAGTTACCCGTCCTGTA	4432
Db	4351	TAATGGCGGTATTTCAGAGCCTCGCCCCCATTTGGCAACCAATACCTGACTCGTAATCTGTA	4410
OY	4433	ATTAAGTGTAAATCAATAAACCGGTGATTCGTTTCAGTTGAAGTTTGTCCTCCTGTCCT	4492
Db	4411	ATTGCTTGTTAATCAATAAACCGTTTAATTCGTTGACGTTGAAGTTTGTCCTCCTGTCAT	4470
OY	4493	TCTTATCTTATCGGTTACCAATGGTTATAGCTTACACATTAAGTCTTGCTGCGCTTGC	4552
Db	4471	TTCCTTCTTATCTAGTTTCATGGCTACGTAGATTAATTAGCATGGCGGTTAATCAATTAA	4530
OY	4553	GATAAAAGACTTACGTCATCGGGTTACCCCTAGTATGAGTTGCCCACTCCCTCTGTC	4612
Db	4531	CTACAAGGA-----ACCCCTAGTATGAGTTGGCCCACTCCCTCTGTC	4574
OY	4613	GCGCTCGCTCGCTCGGTGGGCGCTGCGGACCAAAAGTCCGACAGACGGCAGAGCTCTGCTC	4672
Db	4575	GCGCTCGCTCGCTCACTAGAGCGCGGGGACCAAAAGTCCGCGCCGAGCGGCTTGGCC	4634
OY	4673	TGCGCGGCCACCGAGCGGACGCGCGCAGAGAGGGAGTGGGCCAA 4718	
Db	4635	GCGCGCGCTCAGTAGCGGACGCGCGCGCAGAGAGGGAGTGGGCCAA 4680	

RESULT 5  
US-10-163-886-1  
; Sequence 1, Application US/10163886  
; Publication No. US20020187129A1  
; GENERAL INFORMATION:  
; APPLICANT: Johnson, Philip R.  
; TITLE OF INVENTION: Adeno-Associated Virus Materials and  
; Methods  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 6300 Sears Tower, 233 S. Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/163,886  
; FILING DATE: 04-Jun-2002  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/292,703  
; FILING DATE: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: No. US20020187129A1and, Greta E.  
; REGISTRATION NUMBER: 35,302  
; REFERENCE/DOCKET NUMBER: 31975  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (312) 474-6300

```

; TELEFAX: (312) 474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4680 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-10-163-886-1

```

Query Match	65.3%	Score 3082.4;	DB 15;	Length 4680;
Best Local Similarity	80.0%;	Pred. No. 0;		
Matches 3781; Conservative	0;	Mismatches 891;	Indels 54;	Gaps 11;

QY	1	TTGCCACTCCCTCTCTGCGCGCTTCGCTCGCTCGGTGGGGCTGCCGACCAAGTCCGC	60
Db	1	TTGGCCACTCCCTCTCTGCGCGCTTCGCTCGCTCACTGAGGCCGGGCGACCAAGTCCGC	60
QY	61	AGACGGCAGAGCTCTGCTCTGCGCGCCCAACCGAGCGAGCGCGCAGAGGGAGTG	120
Db	61	CGACGCCCCGGCTTGGCCCGGGCGCTCTAGTGAGCGAGCGCGCAGAGAGGAGTG	120
QY	121	GGCAACTCCATCACTAGGGGTAAATCGCAAGCGCCTCCACGCTGCCGCTCAGCCTGA	180
Db	121	GCCAATCCATCACTAGGGGT-----TCTGAGGGGTGAGTCTGTGA	163
QY	181	CGTAATTACGTATAGG---GAGTGTCTCTGATATAGCTGTCACTGAGTCTTTTC	237
Db	164	CGTGAATTACGTATAGGTTAGGGAGGTCTGTATTAGAGTCACTGAGTG-TTTTGC	222
QY	238	GACATTTTGGACACCACTGTGCCATTTAAGGTATATATGCGCAGTGAGCGACAGAT	297
Db	223	GACATTTTGGACACCACTGTGTCACTGTGGTATTAAAGCCGAGTGAGCAGCAGGGT	282
QY	298	CTCCATTTTG-ACCGCGAAATTGAACGAGCAGCAGCCATGCCGGCTTCTACGAGATCG	356
Db	283	CTCCATTTTGAAGCGGAGGTTGAACGCGCAGCCGCGCATGCCGGGTTTACGAGATTG	342
QY	357	TGATCAAGGTGCGAGCGACTGGACGACACCTGCCGGCATTTCTGACTGTTGTGA	416
Db	343	TGATTAAGTCCCCAGCGACTTGAACGGGCATCTGCCCGCATTTCTGACAGCTTTGTGA	402
QY	417	GCTGGGTGCCGAGAAAGAAATGGGAGCTGCCCGGATTTGACATGATCTGAATCTGA	476
Db	403	ACTGGGTGCCGAGAAAGAAATGGGAGTTGCCCGCATTTGACATGATCTGAATCTGA	462
QY	477	TTGAGCAGGACCCCTGACCGTGGCCGAGAAGCTGCAGCGCATTCCTGTCATATGCG	536
Db	463	TTGAGCAGGACCCCTGACCGTGGCCGAGAAGCTGCAGCGCATTTCTGACGGAATGCG	522
QY	537	GCCGCGTAGTAAGCCCCCGAGGCCCTCTTTTGTTCAGTTCGAGAAGGGCGAGTCT	596
Db	523	GCCGCTGAGTAAGCCCCCGAGGCCCTTTCTTTGTGCAATTTGAGAAGGAGAGAGCT	582
QY	597	ACTTCCACTCCATATTCTGTGTGAGAACCAACGCGGGGTCAATCCATGTGCTGGCCGCT	656
Db	583	ACTTCCACATGACGCTGCTGTGGAACCAACCGGGGTGAATCCATGTGTTTGGGACGTT	642
QY	657	TCCTGAGTCAGATTAGGACAACTGTGTGAGACCATCTACCGCGGATCGAGCCGACCC	716
Db	643	TCCTGAGTCAGATTGCGGAAAACTGATTCAGAGAAATTACCGCGGATCGAGCCGACTT	702
QY	717	TGCCCAACTGTTGCGCGTGAACCAAGCGGTAATGCGCGGAGGGGGAACAAGGTGG	776
Db	703	TGCCAAACTGTTGCGCGTCAACAAGAACCAGAAATGCGCGGAGGGGGAACAAGGTGG	762
QY	777	TGACAGAGTGTACATCCCAACTACCTCTGCCCCAAGCTCAGCCCGAGCTGCAAGTGG	836
Db	763	TGGATGAGTGTACATCCCAATTACTTGTCTCCCAAAAACCCAGCTTGAGCTCCAGTGGG	822
QY	837	CGTGACTAACATGAGAGATATAATAGCGCCTGTTTGAACCTGCGCGAGCGCAACGCGC	896

|||||  
Db 823 CGTGACCTAATATGGAACAGATATTAAAGCCGCTGTTGAATCTCACGGAGCGTAAACGGT 882  
QY 897 TCGTGGCGCAGACCTGACCACGTCAGCCAGACCAGAGCAGAGCAAGAGAAATCTGA 956  
Db 883 TGGTGGCGCAGCATCTGACGCACGCTGTGCGAGACGCAGAGCAGAAACAAGAAATCAGA 942  
QY 957 ACCCCAATTCTGACGCGCCTGTCAATCCGCTCAAAAACTCCGCGCGCTACATGAGAGTGG 1016  
Db 943 ATCCCAATTCTGATGCGCCGCTGATCAGTCAAAAACTTCAGCCAGTACATGAGAGTGG 1002  
QY 1017 TCGGTGGCTGTGGAACCGGGCATCACCTCGAGAAAGCAGTGGATCCAGAGGAGCAGG 1076  
Db 1003 TCGGTGGCTGTGGAACAAGGGGATTAATCTCGGAGAAAGCAGTGGATCCAGAGGAGCAGG 1062  
QY 1077 CCTGTACATCTCTTCAACGCGCTTCAACTCGCGGTCCAGATCAAGCGCTCTGG 1136  
Db 1063 CCTCATACATCTCCTTCATGCGGCTCCAATCGCGGTCCCAATCAAGGCTGCCCTGG 1122  
QY 1137 ACAATGCCGGAAGATCATGCGCTGACCAAAATCCGCGCCGACTACCTGTAGGCCCG 1196  
Db 1123 ACAATGCCGGAAGATTAATGAGCCTGACTTAAACCGCCCCGACTACCTGTGGCCAGC 1182  
QY 1197 CTCCGCGCGCGGACATTAACCAACCGCATCTACCGCATCTTGAGCTGAACGGCTAGC 1256  
Db 1183 AGCCCGTGAAGACATTTCCAGCAATCGATTTATAAATTTGAACTTAAACGGGTACG 1242  
QY 1257 AACCTGCGTACGCGCGGCTCCGCTTTCTCGGCTGGGCCCAAGAAAGTTGCGGAAGCGCA 1316  
Db 1243 ATCCCAATATGCGGCTTCCGCTTTCTGGGATGGGCCACGAAAAAGTTGCGCAAGAGA 1302  
QY 1317 ACACCATCTGGCTGTGTTGGGCGGCAACCAAGGCAAGCAACATCGCGGAAGCCATCG 1376  
Db 1303 ACACCATCTGGCTGTGTTGGGCGCTGCAACTACCGGAAGAACCAACATCGCGGAGCCATAG 1362  
QY 1377 CCCACGCGTCCCTTTCTACGCGCTGCTCAATGGAACCAATGAGAACTTCCCTTCAATG 1436  
Db 1363 CCCACACTGTGCCCTTCTACGCGTCCGTAACAGCAACATGAGAACTTCCCTTCAACG 1422  
QY 1437 ATTGCGTCGACAGATGATGATCTGTGGGAGGAGGCAAGATGACGGCCAAGTCTGG 1496  
Db 1423 ACTGTGTGACAGATGATGATCTGTGGGAGGAGGGAAGATGACCCCAAGTCTGTGG 1482  
QY 1497 AGTCCGCCAAGGCCATTTCTCGGCGGACAGAGGTGCGGCTGGAACCAAAAGTGCAGTCT 1556  
Db 1483 AGTCCGCCAAGGCCATTTCTCGGAGGAAGCAAGGTGCGGCTGGAACCAAAATGCAAGTCT 1542  
QY 1557 CCGCCACAGATGACCCCAACCCCGTGAATGTCACCTCCAACCAACATGTGCGCGTGA 1616  
Db 1543 CCGCCACAGATGACCCCACTCCCGTGAATGTCACCTCCAACCAACATGTGCGCGTGA 1602  
QY 1617 TTGACGGGAACGACCAACCTTCGAGCACCAACCGCTTGACAGACCGGATGTTCAAT 1676  
Db 1603 TTGACGGGAACCTCAACGACCTTCGAACACCAACCGCTTGCAAGACCGGATGTTCAAT 1662  
QY 1677 TTGAATCAACCGCGCTGTGAGCATGACTTTGGCAAGGTGACAAAGCAAGAAAGTCAAG 1736  
Db 1663 TTGAATCAACCGCGCTGTGATCATGACTTTGGGAAGTCAACCAAGCAAGAAAGTCAAG 1722  
QY 1737 AGTCTTCCGCTGGGCGCAGGATCACTGTGACCGGAGTGGCGCATGATTTCTACGTGAG 1796  
Db 1723 ACTTTTCCGCTGGGCGCAAGGATCACTGTGAGGTGAGCATGATTTCTACGTCAAAA 1782  
QY 1797 AGGTGAGCCAACAAAAGACCCGCCCGCATGACGGGATTAAGCAAGCCCAAGCGGG 1856  
Db 1783 AGGTGAGCCAAAGAAAGACCCGCCCGCATGACGGGATTAAGTGAAGCCCAAGCGGG 1842  
QY 1857 CCTGCCCTCAGTCCGGATTCATGACGCTCAGACGCGGAAGAGCTCCGGTGAATTTG 1916  
Db 1843 TCGCGAGTCAGTTCGCGCAGCCATCGACGTCAGACGCGGA---AGCTTCGATCAACTACG 1899  
QY 1917 CCGACAGGTACCAAAACAATGTTCTGTCACGCGGGCATGCTTCAGATGCTGTTCCCT 1976  
|||||

Db 1900 CAGACAGGTACCAAAACAATGTTCTGTCACGCTGGGCAATCTGATGCTGTTCCCT 1959  
QY 1977 GCAAGCATGCGAGAGAATGAATCAGAAATTTCAACATTTGCTTCAACGACGGGACGAG 2036  
Db 1960 GCAGACAATGCGAGAGAATGAATCAGAAATTTCAATATATCTCTCACTCACGACAGAAAG 2019  
QY 2037 ACTGTCAGAGTGTCTCCCGGCGGTGTCAGAAATCTCAACCGGT---CGTCAGAAAGAGA 2093  
Db 2020 ACTGTTAGAGTGTCTTCC---CGTGCAGAAATCTCAACCGGTTTCTGCTCAAAAAGG 2076  
QY 2094 CGTATCGGAAACTCTGTGCCATTCATCATCTGCTGGGGCGGGCTCCCGAGATGTCTGCT 2153  
Db 2077 CGTATCAGAAACTGTGCTACATTCATCATATCATGGAAGAGTGCCAGA---CGCTGCA 2133  
QY 2154 CCGGCTCGCATCTGTGTCACACGTGACCTGGATGACTGTGTTCTGAGCAATAATGACTT 2213  
Db 2134 CTGCTCGCATCTGTCAATGTGGAATTTGGATGACTGCACTTTGAAACAATAATGATTT 2193  
QY 2214 AAACCAAGTATGCTGCCGATGTTATCTTCCAGATTGGCTGAGGACAACTCTGTAG 2273  
Db 2194 AAATCAGTATGCTGCGCATGTATCTTCCAGATTGGCTGAGGACACTCTCTGAA 2253  
QY 2274 GGCATTCCGAGTGTGGGACTTGAACCTGAGCCCGCAAGCCCAAGCCCAACAGCAA 2333  
Db 2254 GGAATTAAGACAGTGTGGAAGCTCAAACTGGCCACCAACCAAAAGCCGCAAGCGG 2313  
QY 2334 AAGCAGACGACGCGCGGTCTGTGCTTCTGCTACAACTCGGACCTCGACCTTCAAC 2393  
Db 2314 CATAGGACGACAGCAGAGGGGTCTGTGCTTCTGCTGACAGTAACTCGGACCTTCAAC 2373  
QY 2394 GGAATCGACAAAGGGGAGCCCGTCAACGCGCGGACGACGCGGCCCTCGAGCAGCAAG 2453  
Db 2374 GGAATCGACAAAGGAGAGCCCGTCAACGAGGACAGCCCGCGGCCCTCGAGCAGCAAA 2433  
QY 2454 GCCTACGACCAAGCAGCTCAAAAGCGGGTGACAAATCCGTACCTGCGGTATAACACGCCG 2513  
Db 2434 GCCTACGACCCGCAAGCTCGACAGCGGAGACAAACCCGTACTCAAGTAAACCAACGCCG 2493  
QY 2514 GCCGATTTCAGGACGCTTCGAAGAAAGATACGTCTTTGGGGGCAACCTCGGGCGAGCA 2573  
Db 2494 GCGGATTTCAGGACGCGCTTAAAGAAAGATACGTCTTTGGGGGCAACCTCGGACGAGCA 2553  
QY 2574 GTCTTCCAGGCCCAAGAGCGGTTCTCGAACCTCTCGCTGTGTTGAGGAAGGCGTAAAG 2633  
Db 2554 GTCTTCCAGGCCGAAGAAAGAGGTTCTTGAACCTCTGGGCTGTGTTGAGGAACCTGTTAG 2613  
QY 2634 ACGGCTCTGGAAGAAAGAGTCGCGTGAAGCAGTGCACCAAGACCAGACTCCTCTCG 2693  
Db 2614 ACGGCTCCGGGAAAAAAGAGCGCGTGAAGCACTCTCTGTGGAAGCCAGACTCCTCTCG 2673  
QY 2694 GGCATCGGCAAGACAGCGCCAGACCCCGCTAAAAAGAGACTCAATTTGTCAGACTGGC 2753  
Db 2674 GGAACCGGAAGGCGGGCCAGACCTTCGAAGAAAAAGATTGAATTTGTCAGACTGGA 2733  
QY 2754 GACTCAGATCAGTCCCGCATCCACAACCTCTCGAGAACTCCAGCAACCCCGCTGCT 2813  
Db 2734 GACGCAAGACTCAGTACCTGACCCCAAGCCTCTCGGACAGCCACAGACCCCTCTGCT 2793  
QY 2814 GTGGACCTACTACAATGGCTTCAGGCGGTGGCGCACCAATGGCAGACAATAACGAAGGC 2873  
Db 2794 CTGGGAATTAATACGATGGCTACAGGCAGTGGCGCACCAATGGCAGACAATAACGAGGC 2853  
QY 2874 GCCGACGAGTGGGTAATGCTTCAGGAATTTGGCATTTGCCATGCGTGGCGAC 2933  
Db 2854 GCCGACGAGTGGGTAATTCCTCGGAATTTGGCATTTGCCATGATGAGGCGAC 2913  
QY 2934 AGAGTCATCACCAACGACGACCCGAACCTGGGCCCTGCCACTCAACAACCAACCTTAC 2993  
Db 2914 AGAGTCATCACCAACGACGACCCGAACCTGGGCCCTGCCACTCAACAACCAACCTTAC 2973  
QY 2994 AAGCAATCTCCAGTGTCTTCAACGGGGGCCAGCAACGACAACCACTACTTGGCTACAGC 3053  
Db 2974 AAACAAATTTCCAGCCAATCA---GAGGCTCGAACGACAATCACTACTTGGCTACAGC 3030  
|||||

QY 3054 ACCCCCTGGGGTATTTTGATTTCACAGATTCACCTGCCACTTTTCAACCACTGACTGG 3113  
Db 3031 ACCCCTGGGGTATTTTGACTTCAACAGATTCACCTGCCACTTTTCAACCACTGACTGG 3090  
QY 3114 CAGCGACTCATACAACAATTTGGGATTCGGGCCCCAAGAGACTCAACTTCAACTCTTC 3173  
Db 3091 CAAGAATCTCATACAACAACCTGGGATTCGACCCCAAGAGACTCAACTTCAAGCTCTTT 3150  
QY 3174 AACATCCAAGTCAAGGAGGTACCGACCAATGATGGCGTCAACAACCTGCTAATAACTT 3233  
Db 3151 AACATTCAGTCAAGAGGTACCGCAAGATGACGGTACGACGACGATTCGAATTAACCTT 3210  
QY 3234 ACCAGCAGGTTCAAGTCTTCTCGACTCGAGTACCAAGCTTCGTAAGTCTCGGCTCT 3293  
Db 3211 ACCAGCAGGTTCAAGTCTTACTGACTCGAGTACCAAGCTTCGTAAGTCTCGGCTCG 3270  
QY 3294 GCGCACCAGGCTGCTCCCTCCGTTCCGGCGGAGCGTGTTCATGATTCGCAATACGGC 3353  
Db 3271 GCGCATCAAGGATGCTCCCGCGTCCGACAGACGCTTTCATGATGCGCACAGTATGA 3330  
QY 3354 TACCTGACGCTCAACAATGGCAGCCAGCCGTTGGAGCGTTCATCTTTTACTGCTGGA 3413  
Db 3331 TACCTCAGCCTGAACAACGGGAGTACGGCAGTAGGACGCTCTTCATTTTACTGCTGAG 3390  
QY 3414 TATTTCCCTTCTCAGATGCTGAGAACGGGCAACAATTACCTTACCTTACCTTTGAG 3473  
Db 3391 TACTTTCTTCTCAGATGCTGCGTACCGGAAACAATTACCTTACCTTACCTTTGAG 3450  
QY 3474 GAAGTCCCTTTCACAGCAGCTACCGCGCAGCAGAGCCGTGAGCCGCTGATGATCCT 3533  
Db 3451 GAGTTCCTTTCACAGCAGCTACCGCTCAGCAGAGTCTGAGCCGCTCATGATCCT 3510  
QY 3534 CTATGACCAATACCTGTATTACCTGAACAAGTCAAAATCAATCCGGAAGTCCCA 3593  
Db 3511 CTATGACCAATACCTGTATTACCTGAACAAGTCAAAATCAATCCGGAAGTCCCA 3570  
QY 3594 AACCAAGACTTGTGTTTACCCGCTGCTCCAGCTGCGATGCTGTTTCAAGCCCAAAAC 3653  
Db 3571 CAGTCAAGGCTTCAAGTTTCTCAGGCGCGGAGCGAGTGAATTCGGGACCACTTGAAGAC 3630  
QY 3654 TGGCTACCTGACCCCTGTATTCGGCAGCAGCCGCTTCTAACAACAACAAGACCAAC 3713  
Db 3631 TGGCTTCTGACCCCTGTATTCGGCAGCAGCGAGTATCAAGACACTCTCGATTAACAC 3690  
QY 3714 AACAGCAATTTTACCTGAGCTGTGCTTCAAAATATACTCAATGGGCTGAATCCATC 3773  
Db 3691 AACAGTGAATACTGTGAGCTGAGCTCAAGTACCACTCAATGGCAGAGACTCTCTG 3750  
QY 3774 ATCAACCTGGCACTGTATGCTTCAACAAGAGCAGCAAGCAAGTCTTCTCCATG 3833  
Db 3751 GTGAATCGGGGGCCCGCATGCAAGCACAAGAGCATGAAGAAATTTTCTCCAG 3810  
QY 3834 AGCGGTTCATGATTTTGGAAAAGAGAGCGCGGAGCTTCAACATGCTGATGGACAAT 3893  
Db 3811 AGCGGGGTCTCATCTTTGGGAAGCAGGCTCAGAGAAAACAAATGTGAACATTTGAAG 3870  
QY 3894 GTCATGATTAACAAGAGGAATTAAGCCACTAACCTTGGGCCACCGAAAGATTT 3953  
Db 3871 GTCATGATTAACAAGAGGAATTCGAACAACCAATCCGCTGCTACGAGCAGTAT 3930  
QY 3954 GGGACCGTGGCAGTCAATTTCCAGAGCAGCAGCAGACAGACCTGCGACCGGAGATGTCAT 4013  
Db 3931 GGTTCTGTATCTACCAACCTCCAGAGGCAACAGACAAAGCAGCTACCGCAGATGTCAAC 3990  
QY 4014 GCTATGAGCATTAACCTGCGATGCTGTGCAAGATAGAGAGCTGACCTGAGGATGCC 4073  
Db 3991 ACACAAGGCGTCTTCCAGGCGATGCTGCGCAGAGCAGAGATGTACTTCAAGGCGCC 4050  
QY 4074 ATTTGGCCAAATTCCTCACAACAAGAGCACTTCAACCGCTCTCTTATGAGCGGC 4133  
Db 4051 ATCTGGCAAAAGATTCACACACGAGAGCAATTTCAACCCCTCTCCCTCATGGGTGA 4110

QY 4134 TTGGAATCAAGAACCCGCTCTCAGATCCTCATCAAAAAACGCTGTTCTGCGAAT 4193  
Db 4111 TTGGAATTAACAACCCCTCTCAGATTTCTCATCAAGAACCCGCTGACTGCGAAT 4170  
QY 4194 CCTCGGCGGAGTTTTCAGCTACAAGTTTGCTTCTCATTCATCAACCAATACTCCACAGA 4253  
Db 4171 CCTGCAACCACTTCAAGTGGCGCAAGTTTGCTTCTCATCAACAGTACTCCACGGA 4230  
QY 4254 CA-AGTGAAGTGAATTTGAATGGAGCTGCAGAAAGAAAACAGCAAGCGCTGGAATCC 4312  
Db 4231 CACGTCAGCGTGAAGATCGAGTGGAGCTGCAGAAAGAAAACAGCAACGCTGGAATCC 4290  
QY 4313 CGAAGTGAATCAACATCCAATTAATGCAAAATCTGCCAAGCTGATTTTACTGTGACAA 4372  
Db 4291 CGAATTCAGTACACTTCCAATCAACAAGTCTGTTAATCTGGAATTAACGCTGATAC 4350  
QY 4373 CAATGACTTATTAATCAATTAACCGGTTGATTCGTTCAATGAACTTTGCTCTCTGCT 4432  
Db 4351 TAATGGCGTGTATTCAGAGCCTGCGCCCATTTGGCACCAGATACCTGTAATCTGTA 4410  
QY 4433 ATTAGTGTATTAATCAATTAACCGGTTGATTCGTTCAATGAACTTTGCTCTCTGCT 4492  
Db 4411 ATTAGTGTATTAATCAATTAACCGGTTGATTCGTTCAATGAACTTTGCTCTCTGCT 4470  
QY 4493 TCTTATCTTAATCGGTTACCATGTTTATAGCTTACACATTAATGCTTGTGCTGCT 4552  
Db 4471 TCTTATCTTAATCGGTTACCATGTTTATAGCTTACACATTAATGCTTGTGCTGCT 4530  
QY 4553 GATTAAGACTTACGTCATCGGTTTACCCCTAGTATGATGAGTTGCCACTCCCTCTGCT 4612  
Db 4531 CTACAAGG-----ACCCTAGTATGAGTTGGCCAATCCCTCTCTGCT 4574  
QY 4613 GCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4672  
Db 4575 GCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4634  
QY 4673 TGCGGCCCCACCGAGCGAGCGCGCGCAGAGAGGAGTGGCAA 4718  
Db 4635 GGGCGGCTCAGTGAGCGAGCGAGCGCGCAGAGAGGAGTGGCAA 4680

RESULT 6  
US-10-263-127-1  
; Sequence 1, Application US/10263127  
; Publication No. US20030082145A1  
; GENERAL INFORMATION:  
; APPLICANT: Johnson, Philip R.  
; TITLE OF INVENTION: Adeno-Associated Virus Materials and Methods  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 6300 Sears Tower, 233 S. Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/263,127  
; FILING DATE: 02-Oct-2002  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/10/077,294  
; FILING DATE: 15-Feb-2002  
; APPLICATION NUMBER: 09/691,604  
; FILING DATE: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: No. US20030082145A1and, Greta E.



REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 31975  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 474-6300  
TELEFAX: (312) 474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4680 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-10-263-127-1

Query Match 65.3%; Score 3082.4; DB 15; Length 4680;  
Best Local Similarity 80.0%; Pred. No. 0;  
Matches 3781; Conservative 0; Mismatches 891; Indels 54; Gaps 11;

QY 1 TTGCCACTCCCTCTCTGCGCGCTCGCTCGGTGGGCTTGGCAAAAGTCCGC 60  
DB 1 TTGGCCACTCCCTCTCTGCGCGCTCGCTCGGTGGGCTTGGCAAAAGTCCGC 60  
QY 61 AGACGGCAGAGCTCTGCTCTGCGCGCGCCACGAGCGAGCGCGCAGAGGGAGTG 120  
DB 61 CGACGCGCGGCTTGGCGCGCGCGCTCAGTGAAGCGAGCGCGCAGAGGGAGTG 120  
QY 121 GGCACTCCATCACTAGGGGTATCGCGAAGCGCCTCCACGCTGCGCGTCAAGCTGA 180  
DB 121 GCCAATCCATCACTAGGGGTATCGCGAAGCGCCTCCACGCTGCGCGTCAAGCTGA 180  
QY 181 CGTAATTAAGTCAATAGGG--GAGTGTCTGTATTAAGTGTACAGTGTCTTTTGC 237  
DB 164 CGTGAATTACGTATAGGGTTAGGAGGTCTGTATTAAGAGTCAAGTGTG-TTTTGC 222  
QY 238 GACATTTTGCAGACCACTGCGCATTTAGGTATATATGCGGAGTGAAGCAGAGAT 297  
DB 223 GACATTTTGCAGACCACTGCGCATTTAGGTATATATGCGGAGTGAAGCAGAGAT 282  
QY 298 CTCCATTTTG-AACGCGAAATTTGAACGAGCAGCAGCATGCGGCTTCTACGAGATCG 356  
DB 283 CTCCATTTTGAAGCGGAGGTTGAACGCGCAGCAGCATGCGGCTTCTACGAGATCG 342  
QY 357 TGATCAAGGTGCGAGCGAAGCTGAGCAGCAGCCTGCGGCAATTTCTGACTGTTTGA 416  
DB 343 TGATTAAGTCCCCAGCGAAGCTGAGCAGCAGCCTGCGGCAATTTCTGACTGTTTGA 402  
QY 417 GCTGGTGGCGGAGAGGAATGGAGTGGCGCGGATTTCTGACATGATCTGAATCTGA 476  
DB 403 ACTGGTGGCGGAGAGGAATGGAGTGGCGCGGATTTCTGACATGATCTGAATCTGA 462  
QY 477 TTGAGCAGGACCCCTGACCGGTGGCGGAGAGCTGACGCGGACTTCTGTGTCATATGGC 536  
DB 463 TTGAGCAGGACCCCTGACCGGTGGCGGAGAGCTGACGCGGACTTCTGTGTCATATGGC 522  
QY 537 GCCGCGTGAAGAGCGCGGAGCGCTCTTTTGTGTTAGTTGAGAGGGCGAGTCTCT 596  
DB 523 GCCGCGTGAAGAGCGCGGAGCGCTCTTTTGTGTTAGTTGAGAGGGCGAGTCTCT 582  
QY 597 ACTTCCACTCCATATTTCTGTGAGAGCAGCGGCGTCAATCCATGCTGCGCGCT 656  
DB 583 ACTTCCACTCCATATTTCTGTGAGAGCAGCGGCGTCAATCCATGCTGCGCGCT 642  
QY 657 TCCTGAGTCAAGTTAGGAGCAAGCTGTGAGACCATCTACCGCGGATGAGCGGACCC 716  
DB 643 TCCTGAGTCAAGTTAGGAGCAAGCTGTGAGACCATCTACCGCGGATGAGCGGACCC 702  
QY 717 TGCCCAACTGTTGCGGCTGACCAAGACGCGTATGCGCGCGGAGGGGGAACAAGTGG 776  
DB 703 TGCCCAACTGTTGCGGCTGACCAAGACGCGTATGCGCGCGGAGGGGGAACAAGTGG 762  
QY 777 TGAGCAGTGTACATCCCCAACTACTCTCTGCGCAAGACTCAGCCCGAGCTGCAAGTGG 836

DB 763 TGATGAGTGTACATCCCCAACTACTCTCTGCGCAAAACCCAGCTGAGCTCCAGTGG 822  
QY 837 CGTGAATAACATGAGAGATATATAAGCGCTGTTGAACCTGGCCGAGCGCAACGGC 896  
DB 823 CGTGAATAATATGAGACAGATATTAAGCGCTGTTGAATCTCACGAGCGTAACGGT 882  
QY 897 TCGTGGCGCAGCAGCTGACCCAGTCAAGCCAGACCCAGAGCAGAAACAGAGATCTGA 956  
DB 883 TGTGGCGCAGCATCTGACCGCAGTGTGCGAGACCGCAGAGCAGAAACAGAGATCTGA 942  
QY 957 ACCCAATTCTGACGCGCGCTGTCAATCCGTTCAAAAACCTCCGCGCTACATGAGCTGG 1016  
DB 943 ATCCCAATTCTGATGCGCGCGTGTCAATCAAAAACCTTCAGCCAGTACATGAGCTGG 1002  
QY 1017 TCGGCTGCTGTGAGACCGGCGATCACTCCGAGAAAGCAGTGAATCCAGAGACCAAG 1076  
DB 1003 TCGGCTGCTGTGAGACCGGCGATCACTCCGAGAAAGCAGTGAATCCAGAGACCAAG 1062  
QY 1077 CCTGTACATCTCTTCAACGCGCTTCCAACTGCGCGTCCAGATCAAGCGCTCTGG 1136  
DB 1063 CCTGTACATCTCTTCAATGCGCGCTTCCAACTGCGCGTCCAAATCAAGCTGCTTGG 1122  
QY 1137 ACAATGCGGCAAGATCATGCGCGTGAACCAATCCGCGCGCACTACCTGTAGGCCCG 1196  
DB 1123 ACAATGCGGCAAGATTAATGAGCTGAATAAACCGCGCGCACTACCTGTAGGCCCG 1182  
QY 1197 CTCCGCGCGGACATTAACCAACCGCATCTACCGCATCTGAGCTGAACGCGTACG 1256  
DB 1183 AGCCGCGGAGACATTTCCAGCAATCGGATTTATAAATTTTGAACCTAAACGCGTACG 1242  
QY 1257 AACCTGCTACGCGCGCTCTCTTCTGCGCTGGCGCCAGAAAGGTTGCGGAAGCGCA 1316  
DB 1243 ATCCCAATATGCGCGCTCTCTTCTGCGATGGCGCCAGAAAGGTTGCGGAAGCGCA 1302  
QY 1317 ACACATCTGCTGTTTGGCGCGCAACACGCGCAAGCAACATCGCGGAAGCCATCG 1376  
DB 1303 ACACATCTGCTGTTTGGCGCGCAACACGCGCAAGCAACATCGCGGAAGCCATCG 1362  
QY 1437 ATTGCGTGAACAAGATGATCTGTGTTGGAGGAGGCGCAAGATGACCGCCAGAGTCTGG 1496  
DB 1423 ACTGTGTGAACAAGATGATCTGTGTTGGAGGAGGCGCAAGATGACCGCCAGAGTCTGG 1482  
QY 1497 AGTCCGCAAGGCGATTTCTGCGGCGCAGCAAGGTGCGCGTGAACCAAGTGAAGTCTGT 1556  
DB 1483 AGTCCGCAAGGCGATTTCTGCGGCGCAGCAAGGTGCGCGTGAACCAAGTGAAGTCTGT 1542  
QY 1557 CCGCCGATGACCCCGCAGTGTGATGCTGATGCTCACTCAACCAACATGTGCGCGCTGA 1616  
DB 1543 CCGCCGATGACCCCGCAGTGTGATGCTGATGCTCACTCAACCAACATGTGCGCGCTGA 1602  
QY 1617 TTGACGGGAACAGCAGCACTTTCAGAGCAGCAGCGGCTGAGAGCCGATGTTCAAT 1676  
DB 1603 TTGACGGGAACAGCAGCACTTTCAGAGCAGCAGCGGCTGAGAGCCGATGTTCAAT 1662  
QY 1677 TTGAATCAACCCCGCTGAGAGTGAATTTGGCAAGTGAACAAGCAGAGTCAAG 1736  
DB 1663 TTGAATCAACCCCGCTGAGAGTGAATTTGGCAAGTGAACAAGCAGAGTCAAG 1722  
QY 1737 AGTTCTTCGCTGGGCGCAGATCACTGACCGAGGTGGCGCATGAGTTCTACGTCAAG 1796  
DB 1723 AGTTCTTCGCTGGGCGCAGATCACTGAGTTGAGGTGAGCATGAATTTCTACGTCAAG 1782  
QY 1797 AGGTGAGCCAAACAAGAGCCCGCGCGATGACGCGGATTAAGCGAGCCCAAGCGGG 1856  
DB 1783 AGGTGAGCCAAACAAGAGCCCGCGCGATGACGCGGATTAAGCGAGCCCAAGCGGG 1842  
QY 1857 CCTGCCCTCAGTGGCGATCACTGAGCTGAGACGCGGAGAGAGCTCCGCTGACTTTG 1916

Db 1843 TGCGGAGTCAAGTTCGCGACCATCGACGTGACAGCGGA---AGCTTCGATCAACTACG 1899  
Qy 1917 CCGACAGGTACCAAAACAAATGTTCTGTCACGCGGCATGCTTCAGATGCTGTTCCCT 1976  
Db 1900 CAGACAGGTACCAAAACAAATGTTCTGTCACGTCGGCATGATGATGCTGTTCCCT 1959  
Qy 1977 GCAAGACATGCGAGAGATGATGATTAATTTCAACATTTGCTTCACGCACGGGACGAGAG 2036  
Db 1960 GCAGACATGCGAGAGATGATGATTAATTTCAATATATCTGCTTCACTCACGGACAGAAAG 2019  
Qy 2037 ACTGTTCAAGTGTCTTCCCGCGCTGTGAGAACTCAACCCGT---CGTCAGAAAGAGA 2093  
Db 2020 ACTGTTAGAGTGTCTTCC---CGTTCAGAAATCTCAACCCGTTTCTGTCATAAAAGG 2076  
Qy 2094 CGTATCGAAACTCTGTGCCATTCATCTGCTGGGCGGCTCCGAGATTGCTTGTCT 2153  
Db 2077 CGTATCAGAACTGTGTACATTCATCAATGAGAAAGGTGCCAGA---CGCTTGCA 2133  
Qy 2154 CGGCTGCGATCTGTCAACGTGACCTGATGATGATGATGATTTCTGACCAATAATGACTT 2213  
Db 2134 CTGCTGCGATCTGTCAATGTGATTTGATGATGATGATGATCTTGAACAATAATGATTT 2193  
Qy 2214 AAACCAAGTATGCTGCGGATGTTATCTTCCAGATTGGCTCGAGACAACTCTCTGAG 2273  
Db 2194 AAATCAGTATGCTGCGGATGTTATCTTCCAGATTGGCTCGAGACAACTCTCTGAA 2253  
Qy 2274 GGCATTCGCGAGTGTGTGGACTTGAACCTTGAGGCCCCGAAAGCCCAAGCAACGAA 2333  
Db 2254 GGAATAGACAGTGTGTGAAGCTCAACCTTGCCCCACCAACCAAGCCCGCAGAGCGG 2313  
Qy 2334 AAGCAGACGACGCGCGGGGTCTGTGCTTCTGCTGCTACAAAGTACCTCGAACCTTCAAC 2393  
Db 2314 CATAGGACGACGACGAGGGTCTGTGCTTCTGCTGCTACAAAGTACCTCGAACCTTCAAC 2373  
Qy 2394 GGAATCGACAGGGGAGGAGCCCGTCAACGCGGCGGACGACGCGGCTTCGAGACGACAA 2453  
Db 2374 GGAATCGACAGGGGAGGAGCCGCTCAACGAGGCGACGCGGCGGCTTCGAGACGACAA 2433  
Qy 2454 GCCTACGACGACGAGCTCAAGCGGGGTGAATCCGTACCTGCGGTATTAACACGCGGAC 2513  
Db 2434 GCCTACGACGCGGACGCTCGACAGCGGACCAACCGTACCTCAAGTACAACACGCGGAC 2493  
Qy 2514 GCGGAGTTTCAAGAGCGCTCTGCAAGAAATACGTCTTTTGGGGCAACCTCGGGGAGCA 2573  
Db 2494 GCGGAGTTTCAAGAGCGCTTAAAGAAATACGTCTTTTGGGGGCAACCTCGGACGAGCA 2553  
Qy 2574 GTCTTCAAGCCAAAGAGCGGTTCTGAACTCTCGGTCTGTTGAGGAAAGCGCTAAG 2633  
Db 2554 GTCTTCAAGCGAAAGAGAGGTTCTTGAACCTCTGGGCTGTTGAGGAACTGTTAAG 2613  
Qy 2634 ACGGCTCTGGAAGAAAGAGTCCGCTGAGACAGTCCGACCAAGAGCCAGACTCCTCTCG 2693  
Db 2614 ACGGCTCCGGGAAAAAGAGCGCGGTAGCACTCTCTGTGAGCCAGACTCCTCTCG 2673  
Qy 2694 GGCATCGGCAAGACAGCGCAGCGCGCTAAAGAGACTCAATTTTGGTCAAGCTGGC 2753  
Db 2674 GGAACCGGAAGCGCGGCGCAGCAGCTGCAAGAAAGATTTGTCAGACTGGA 2733  
Qy 2754 GACTCAAGTCAATCCCCGATCCCAACCTCTCGGAGAACCTTCAGCAACCCCGCTGCT 2813  
Db 2734 GACGCAAGTCAATCTGACCCCGCAGCTCTCGGACAGCCACGACGACCCCTCTGCT 2793  
Qy 2814 GTGGGACCTTACATCAATGCTTCAAGCGGTGGCGCACCAATGGCAGCAATTAAGAGGC 2873  
Db 2794 CTGGGAATTAATCGATGGCTACAGGAGTGGCGCACCAATGGCAGCAATTAAGAGGC 2853  
Qy 2874 GCGGACGAGTGGGTATGCTCTCAGAAATGGCAATGGCAATTCAGATGCTGGCGAC 2933  
Db 2854 GCGGACGAGTGGGTATGCTCTCGGAAATGGCAATGGCAATTCAGATGCTGGCGAC 2913  
Qy 2934 AGAGTCAATCAACACGACCGCAGACTGGGCTGGGCACTTACCAATTAACACCTCTAC 2993  
Db 2914 AGAGTCAATCAACACGACCGCAGACTGGGCTGGGCACTTACCAATTAACACCTCTAC 2973

Qy 2994 AAGCAATCTCAGTGTCTTCAACGGGGCCAGCAACGACCACTACTTGGCTACAGC 3053  
Db 2974 AAACAAATTTCCAGCCAAATCA---GGAAGCTCGAACGACAAATCACTACTTGGCTACAGC 3030  
Qy 3054 ACCCGCTGGGGTATTTGATTTCAACAGATTCCACTGCCACTTTTCAACAGTACTGG 3113  
Db 3031 ACCCTTGGGGTATTTGATTTCAACAGATTCCACTGCCACTTTTCAACAGTACTGG 3090  
Qy 3114 CAGGACTCATCAACAACAAATGGGGATTCCGGCCCAAGAGACTCAACTTCAACTCTTC 3173  
Db 3091 CAAGACTCATCAACAACAACTGGGGATTCCGACCCCAAGAGACTCAACTTCAAGCTCTT 3150  
Qy 3174 AACATCCAAGTCAAGAGGTCAAGACGATGATGCGCTCAACAACCATCGTAATACTT 3233  
Db 3151 AACATTCAGTCAAGAGGTCAAGACGATGATGCGGTACGACGATTTGCCAATACTT 3210  
Qy 3234 ACCAGCAGGTTCAAGTCTTCTGGAAGTCCGAGTACAGCTTCCGTACGCTCCGCTCT 3293  
Db 3211 ACCAGCAGGTTCAAGTGTATTAAGTCTCGAGTACAGCTCCGTAAGCTCCGCTCTG 3270  
Qy 3294 GCGACACAGGCTGCTCCCTCCGTTCCCGGCGGACGTTGATGATTTCCGCAATACGGC 3353  
Db 3271 GCGCATCAAGGATGCTCCCGCGTTCCAGCAGACGTTTCAATGATGATGATGATGATG 3330  
Qy 3354 TACCTGACGCTCAACAATGGCAAGCCAGCCGTGGAGCTTCACTTTTACTGCTGGA 3413  
Db 3331 TACCTCAACCTGAACAACGGGAGTCAAGGACGATGAGACGCTCTTACTTACTGCTGAG 3390  
Qy 3414 TATTTCCCTTCTCAGATGCTGAGAACGGGCAACACTTACCTTACGTAACACTTTGAG 3473  
Db 3391 TACTTCTCTTCAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3450  
Qy 3474 GAAGTCCCTTTCACAGCAGCTACGCGGACAGCCAGACGCTGGAACCGGCTGATGAATCT 3533  
Db 3451 GAGTTCCTTTCACAGCAGCTACGCTCAACAGCAGTGTGGAACCGTCTCATGAATCT 3510  
Qy 3534 CTCAATGACCAATACCTGTATTAAGTGAACAGAACTCAAAATCAAGTCCGGAAGTCCCA 3593  
Db 3511 CTCAATGACCAATACCTGTATTAAGTGAACAGAACTCAAAATCAAGTCCGGAAGTCCCA 3570  
Qy 3594 AACAGGACTTGTGTTAGCGGTGCTCCAGCTGCGATGCTGTTCAAGCCCAAAAC 3653  
Db 3571 CAGTCAAGCTTCAAGTTTCTCAGCGCGGACGAGTGAATTCGGGACCACTGTAAGAC 3630  
Qy 3654 TGGCTACTGACCTGTTTATGCGCAGACGCGGCTTTCTAATAACAAACAGACAAACAC 3713  
Db 3631 TGGCTCTGGAACCTGTTTATGCGCAGACGCGGATATCAAGACATCTGCGATTAACAC 3690  
Qy 3714 AACAGCAATTTTACCTGAGCTGCTTCAAAATATTAAGTGAATGCGGTGAATCCATC 3773  
Db 3691 AACAGCAATTTTACCTGAGCTGCTTCAAAATATTAAGTGAATGCGGTGAATCCATC 3750  
Qy 3774 ATCAACCTGGAAGTCTGATGCTTCAACAAAGACGACGACCAAGTCTTCCCATG 3833  
Db 3751 GTGAATCCGGGCGCGCATGGAAGCCCAAGGAGATGAAGAAAGTTTTCCTCAG 3810  
Qy 3834 AGCGGTGATGATTTTGAAGAAAGAGCGCGGAGCTTCAAACTGATTTGACAAAT 3893  
Db 3811 AGCGGTGATGATTTTGAAGAAAGAGCGCGGAGCTTCAAACTGATTTGACAAAT 3870  
Qy 3894 GTCAATGATTAACAGAGAGAAATTAAGCCACTAACCTGTGGCCACGAAAGATTT 3953  
Db 3871 GTCAATGATTAACAGAGAGAAATTAAGCCACTAACCTGTGGCCACGAAAGATTT 3930  
Qy 3954 GGGACCGTGGCAGTCAATTTCCAGAGCAGCAGCAGACCTGCGACCGGAGATGTGAT 4013  
Db 3931 GGTTCGTATTAACCACTTCCAGAGAGGCAACAGACAGCACTACCGCAGATGTCAAC 3990  
Qy 4014 GCTATGGAGCATTAACCTGGAAGTGTGCAAGATGAGACGTTACTGACGGGTCC 4073  
Db 3991 ACACAAGGCGTTCTTCCAGGAGTGTGCTGCGACGACAGAGATGTATCACTTACGGGCGCC 4050

QY	4074	ATTGGGCCCCAAATTCCTTACACAGATGGACACTTTCACCCGCTCTCCTTATATGGCGGC	4133
Db	4051	ATCTGGGCAAGATTCACACACGGACGGACATTTTACCCCTCTCCCTCATATGGGTGA	4110
QY	4134	TTTGACTCAAGAACCCGCGCTCTCAGATCCTCATCAAAAACAGCGCTGTTCTCGCAAT	4193
Db	4111	TTGGAAGTTAAACACCCCTCTTCCACAGATTTCTATCAAGAACACCCCGTACTGCGAAT	4170
QY	4194	CCTCCGGCGGAGTTTTCAGCTACAAAGTTTGCTTCATTCATCAACCAATACTCCACAGA	4253
Db	4171	CCTTCGACCACTTCAAGTGGCGCAAGTTTGCTTCTTCAATCACAGTACTCCACGGGA	4230
QY	4254	CA-AGTAGTGTGAATTTGAATGGAGCTGCAGAAAGAAACAGCAAGCGCTGGAATCC	4312
Db	4231	CACGCTCAGCGTGGAGATCGAGTGGGAGCTGCAGAAAGAAACAGCAAGCGTGAATCC	4290
QY	4313	CGAAGTCAGTACACATCCAAATTATGCAAAATCTGCCAACGTTGATTTTACTGTGACAA	4372
Db	4291	CGAAATTCAGTACACTTCCAACTACAAACAAGTCTGTTAATCGTGAAGTTACCGGTGATAC	4350
QY	4373	CAATGACTTATATCTGAGCTCGCCCAATTGGCAACCCGTTAAGTTAACCCTGCTGTA	4432
Db	4351	TAATGCGGTATTCAGAGCCTCGCCCAATTGGCACCAAGATNCTGACTCGTAATCTGTA	4410
QY	4433	ATTACGTGTTAATCAATAAACCGGTGATTCGTTCAATTGAATTTGGTCTCTGTCCT	4492
Db	4411	ATTGCTTGTTAATCAATAAACCGGTTAATTCGTTGAGTTGAATTTGGTCTCTGTCAT	4470
QY	4493	TCTTATCTTATCGGTTACCAATGCTTATAGCTTACACATTAATGCTTGGTGGCTGCG	4552
Db	4471	TCTTCTTATCTAGTTTCCATGCGTACGTAGATTAATTAGCATGGCGGTTAATCATTA	4530
QY	4553	GATAAAGACTTACGTCACTCGGGTTACCCCTAGTATGAGTGGCCCACTCCCTCTGCG	4612
Db	4531	CTACAAGA-----ACCCCTAGTATGAGTGGCCCACTCCCTCTGCG	4574
QY	4613	GCGCTCGCTCGCTCGGTGGGCGCTGCGGACCAAAAGTCCGAGACGGCAGAGCTCTGCTC	4672
Db	4575	GCGCTCGCTCGCTCACTGAGGCGGCGGACCAAAAGTCCGCGCAGACGCCCGGCTTGGCC	4634
QY	4673	TGCCGCGCCCCACCGAGCGGACGCGCGCAGAGAGGGAGTGGCCAA	4718
Db	4635	GCGCGCGCTCACTGAGCGGAGCGGCGCGCAGAGAGGGAGTGGCCAA	4680

Query Match	64.8%;	Score 3055.8;	DB 11,	Length 4675;
Best Local Similarity	79.9%;	Pred. No. 0;		
Matches 3779;	Conservative	0;	Mismatches 887;	Indels 61;
			Gaps	13;
0x	1	TTGCCCACTCCCTCTGTGGCGGCTGCGTGGTGGGCGCTGGCGACCAAGGTCCGC	60	

Db	1	TTGGCCACTCCCTCTCTGCGCGCTCGCTCGTCACTGAGCGCGGCGCAACCAAGTCCGCC	60
QY	61	AGACGGCAGAGCTCTGCTCTGCGCGGCCCAACGAGCGAGCGCGCAGAGGAGTG	120
Db	61	CGACGCCCGGCTTGCCCCGGCGGCTCAGTGAAGCAGCGAGCGCGCAGAGGAGTG	120
QY	121	GGCACTCCATCACTAAGGGTATCGCGAAGCGCTCCACGCTGCCGCTCAGCGCTGA	180
Db	121	GCCCACTCCATCACTAAGGGT-----TCTGGAGGGGTGGAGTCTGTGA	163
QY	181	CGTAAATTACGTCAATAGGG--GAGTGTCTGTATTAAGTGTCACTGAGTCTTTTGC	237
Db	164	CGTGAATTACGTCAATAGGGTTAGGAGGTCTGTATTAAGAGTCACTGAGTG-TTTTGC	222
QY	238	GACATTTTGGCAGACCACTGCGCCATTAGGTATATATGCCCCAGTGAAGCAGGAT	297
Db	223	GACATTTTGGCAGACCACTGCTGCTGAGTATTAAGCCGAGTGAAGCAGCGGT	282
QY	298	CTCCATTTTG-ACCGCGAAATTGAACGAGCAGCAGCCATGCCGGCTTTCAAGATCG	356
Db	283	CTCCATTTTGAAGCGGGAGTTGAACGCGCAGCGCATGCCGGGTTTTCAGAGATTG	342
QY	357	TGATCAAGGTGCCGAGCGCACTTGACGAGCACTGCCGGCATTTCTGACTGCTTTGTGA	416
Db	343	TGATTAAGTCCCGCAGCGCACTTGACGGGCACTGCCGGCATTTCTGACAGCTTTGTGA	402
QY	417	GCTGGTGGCCGAGAAAGAAATGGAGCTGCCCCCGGATTTCTGACATGATCTGAATCTGA	476
Db	403	ACTGGTGGCCGAGAAAGAAATGGAGTTGCCCGCAGATTTCTGACATGATCTGAATCTGA	462
QY	477	TTGAGCAGGCAACCCCTGACCGTGCCGAGAGCTGACGCGCACTTCTGTCCAATGGC	536
Db	463	TTGAGCAGGCAACCCCTGACCGTGCCGAGAGCTGACGCGCACTTCTGACGGAATGGC	522
QY	537	GCCGCTGAGTAAGGCCCGGAGGCCCTTTCTTTGTTCACTTGAAGAGGCGAGTCTT	596
Db	523	GCCGCTGAGTAAGGCCCGGAGGCCCTTTCTTTGTTCACTTGAAGAGGAGAGCT	582
QY	597	ACTTCCACCTCCATATTCTGTGAGAGACCAAGGAGTCAATCCATGTTGTCGGCCGCT	656
Db	583	ACTTCCACATGACAGTGTCTGTGAAACCAACGAGGAGTGAATCCATGTTTGGGACGTT	642
QY	657	TCCTGAGTCAGATTAGGACAAGCTGCTGACACCACTACCGCGGATCGAGCCGACCC	716
Db	643	TCCTGAGTCAGATTGCGGAAAACTGATTCAGAGAAATTAACGCGGATCGAGCCGACTT	702
QY	717	TGCCCAACTGTTGCGGGTGAACCAAGCGGCTATAGCCGCCGAGGGGGGACAAGGTGG	776
Db	703	TGCCCAACTGTTGCGGGTCACAAAAGCACAAATGGCCGCCGAGCGGGACAAGGTGG	762
QY	777	TGAGCAGTGTACATCCCAACTACCTCTGCCAAGACTCAGCCCGAGCTGCAGTGGG	836
Db	763	TGATGAGTGTACATCCCAATTACTGTCTCCCAAAACCAAGCTGAGCTTCAAGTGGG	822
QY	837	CGTGAATAACATGAGAGATATATTAAGCGCTTTGAACCTGCGCGGCAAGCGGCAACGCGC	896
Db	823	CGTGAATAATGAGACAATATTAAGCGCTTTGAATCTCAGGAGCGTAAACGCT	882
QY	897	TGCTGGCGCAGCACTGACCCACGTCAGCCAGACCCAGAGCAGAAACAAGGAATCTGA	956
Db	883	TGCTGGCGCAGCATGTACGCAAGTGTGCGCAGCAGCAGAGCAGACAAAGAGATCAGA	942
QY	957	ACCCCAATTCTGACCGCGCTGTCTATCCGGTCAAAAACTCCGCGCGCTACATGAGCTGG	1016
Db	943	ATCCCAATTCTGATGCGCGGTGATCAGATCAAAAACTTACGCCAGGTACATGAGCTGG	1002
QY	1017	TGCGGTGCTGTGGAACGGGGCATCACTCCGAGAAAGCAGTGATCCAGAGGACCAAG	1076
Db	1003	TGCGGTGCTGTGGAACAAGGGATTAACCTGGAGAAAGCAGTGATCCAGAGGACCAAG	1062
QY	1077	CCTGTACATCTCTTCAACGCCGCTTCCAACTGCGGCTCCAGATCAAGCGCGCTCTGG	1136



Db 1063 CCTCATCATCTCTTCAATGCGGCTTCCAACCTCGCGTCCCAATCAAGGCTGCTTGG 1122  
Qy 1137 ACAATGCCGCAAGATCATGCGCTGACCAATCCGCGCCGACTACCTGTAGGCCCG 1196  
Db 1123 ACAATGGGGGAAGATTATGAGCCCTGATAAACCGCCCGACTACCTGTGGCGAGC 1182  
Qy 1197 CTCCGCGCGCGACATTAAACCAACCGCATCTACCGCATCTGTGAGCTGAACGGCTACG 1256  
Db 1183 AGCCCGTGAAGACATTTCAGCAATCGAATTATATAATTTTGGAATTAACCGGCTACG 1242  
Qy 1257 AACCTGCTTACGCGCGCTCCGCTCTTCTCGGCTGGGCGCCAGAAAAGTTTCGGAAAGCGCA 1316  
Db 1243 ATCCCAATATGCGGCTCCGCTCTTCTTGGGATGGGCCACGAAAAAGTTTCGCAAGAGA 1302  
Qy 1317 ACACCATCTGGCTGTTTGGCGCGGCAACCGGGCAAGACCAATCGCGAAGCCATCG 1376  
Db 1303 ACACCATCTGGCTGTTTGGCGCTGCACTAACCGGGAAGACCAATCGCGAAGCCATAG 1362  
Qy 1377 CCCACGCGTGCCTTCTACGCGCTGCTCACTGACCAATGAGAACTTCCCTTCAATG 1436  
Db 1363 CCCACATGTGCCCTTCTACGCGTGCCTTAACTGACCAATGAGAACTTCCCTTCAACG 1422  
Qy 1437 ATTGCGTGCACAAGATGTGATCTGTGGGAGGAGGCAAGATGACGCGCAAGTGTGG 1496  
Db 1423 ACTGTGTCGACAAGATGTGATCTGTGGGAGGAGGAGGAGATGACCGCAAGTGTGG 1482  
Qy 1497 AGTCCGCGCAAGGCCATTCTCGCGCGCAAGAGTGCCTGACCAAAAAGTGCAGTCTGT 1556  
Db 1483 AGTCCGCGCAAAAGCCATTCTCGGAGGAAGCAAGGTGCGCTGACCAAGAAATGCAGTCTCT 1542  
Qy 1557 CCGCCAGATGACCCCAACCCCGCTGATGTCACCTCCAACACCAATGTGCGCCGTGA 1616  
Db 1543 CCGCCAGATGACCCCAACCCCGCTGATGTCACCTCCAACACCAATGTGCGCCGTGA 1602  
Qy 1617 TTGACGGGAACAGACACCACTTTCAGACCAAGAGCCGTTGACAGACCGGATGTTCAAT 1676  
Db 1603 TTGACGGGAACCTCAACGACCTTTCAGACCAAGAGCCGTTGACAGACCGGATGTTCAAT 1662  
Qy 1677 TTGAACCTCACCCGCGCTGTGAGCATGACTTTGGCAAGGTGACAAAGCAGAGTCAAG 1736  
Db 1663 TTGAACCTCACCCGCGCTGTGATGATGACTTTGGGAAGTCAACCAAGAGTCAAG 1722  
Qy 1737 AGTCTTCCGCTGGGCGCAGGATCACTGACCGAGTGGCGCATGAGTTCTACGTAGAA 1796  
Db 1723 ACTTTTCCGCTGGGCGCAAAAGATCACTGTGTTGAGGTGAGCATGAAATTTCTACGTAA 1782  
Qy 1797 AGGTGAGACCAACAAAGACCCCGCCCGATGACGCGGATAAAAGCAGACCCCAAGCGGG 1856  
Db 1783 AGGTGAGACCAAGAAAGACCCCGCCCGATGACGAGATTAAGTGAACCCCAAGCGGG 1842  
Qy 1857 CCTGCCCCCTCAGTGCAGGATTCATGACGTGACAGCGGGAAGAGCTCCGGTGAATTTG 1916  
Db 1843 TCGCGAGTCAGTGCGCAGCATGACGTGACAGCGCGA---AGCTTCGATCAACTACG 1899  
Qy 1917 CCGACAGGTACCAAAACAATGTTCTGTCACGCGGCGATGCTTCAAGTGTGTTCCCT 1976  
Db 1900 CAGACAGGTACCAAAACAATGTTCTGTCACGTGGCATGAACTGATGTTCCCT 1959  
Qy 1977 GCAGACATGCGAGAGATGAATCAAGATTTCAACATTTGCTTCAACGACGGAAGAG 2036  
Db 1960 GCAGACATGCGAGAGATGAATCAAGATTTCAAAATATCTGCTTCACTCAACGACAGAA 2019  
Qy 2037 ACTGTTCAAGTGTCTTCCCGCGCTGACAAATCTCAACCCGT---CGTCAGAAAGAGA 2093  
Db 2020 ACTGTTTGAAGTCTTCC---CGTTCAGAAATCTCAACCCGTTTCTGTCGTCAAAAAGG 2076  
Qy 2094 CGTATCGAAACTCTGTGCCATTCATCTGCTGGGGCGGCTCCGAGATGCTTGTCT 2153  
Db 2077 CGTATCAGAAACTGTGTACATTCATCATATCATGGGAAGGTGCCAGA---CGCTTGA 2133  
Qy 2154 CGGCTGCGATCTGTCAACGTGACCTGAGTGACTGTGTTTGAAGCAATAATGACTT 2213  
Db 2134 CTGCTCGATCTGTCAATGTGATTTGATGACTGATCTTGAACAATAATGATTT 2193

Qy 2214 AAACAGTATGCTGCCGATGTTATCTTCCAGATTTGGCTCGAGACCAACCTCTCTGAG 2273  
Db 2194 AAATCAGGTATGCTGCCGATGTTATCTTCCAGATTTGGCTCGAGACCACTCTCTGAA 2253  
Qy 2274 GGCATTCCGAGTGTGGGACTTGAACCTTGAGCCCGAAGCCCAAGCCCAACAGCA 2333  
Db 2254 GGAATTAAGACAGTGTGAAGCTCAAACTGGCCACCAACCAAGCCCGCAGAGCGG 2313  
Qy 2334 AAGCAGACGACGCGCGGCTGTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2393  
Db 2314 CATTAAGACGACAGAGGGTCTGTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2373  
Qy 2394 GGACTCGACAAGGGGAGCCCGTCAACGCGCGGACGCGCGCCCTCGAGACG-ACAA 2452  
Db 2374 GGACTCGACAAGGGAGAGCCCGTCAACGAGGACAGCGCGCGCCCTCGAGACGTA 2433  
Qy 2453 GGCCTACGACCAAGCTCAAGCGGGTGAACATCCGTAACCTGCTGCTGCTGCTGCTGCTGCT 2512  
Db 2434 AGCTTACGACCGGACGCTCGACAGCGGACCAACCCGTAACCTCAAGTCAACCAACGCGCA 2493  
Qy 2513 CGCGGAGTTTCAAGAGCGCTTTCAGAAAGATACGCTTTTGGGGGCAACCTCGGCGAGC 2572  
Db 2494 CGCGGAGTTTCAAGAGCGCTTTCAGAAAGATACGCTTTTGGGGGCAACCTCGGCGAGC 2553  
Qy 2573 AGCTTTCAGAGCCCAAGAGCGGTTCTCGAAACCTCTCGCTGTGTTGAGGAAGCGCTAA 2632  
Db 2554 AGCTTTCAGAGCCCAAGAGCGGTTCTTGAACCTCTGCGCTGTGAGGAACCTGTTAA 2613  
Qy 2633 GACGCTCTCTGGAAGAAAGCTCCGGTAGAGCAGTCCGCAACAAGCCAGACTCTCTC 2692  
Db 2614 GACGCTCTCCGGGAAAAAGAGCGCGTAGAGCAGTCTCTGTGAGAGCAGACTCTCTC 2673  
Qy 2693 GGGCATCCGCAAGACAGCGCCAGACCCCGCTAAAAAGACTCAATTTTGGTCAAGTGG 2752  
Db 2674 GGGAAACCGGAAGCGCGCCAGCAGCTTCAAGAAAAAGATTGAATTTTGGTCAAGTGG 2733  
Qy 2753 CGACTCAGAGTCACTCCCGATTCACCAACCTCTCGAGAACTTCCAGCAACCCCGCTGC 2812  
Db 2734 AGACGAGACTCAGTACCTGACCCCAAGCTCTCGGACGCCACAGCAGCCCTCTGCG 2793  
Qy 2813 TGTGGACTTACTCAATGCTTCAAGCGGTGCGCAACAATGGCAGACAATTAACGAAG 2872  
Db 2794 TGTGGAACTAATACGATGGCTACAGCAGTGGCGCACCAATGGCAGACAATTAACGAAG 2853  
Qy 2873 CGCGACGAGTGGTAATGCTTCAAGAAATTGGCATTTCCACATGCTGGCGCA 2932  
Db 2854 CGCGACGAGTGGTAATGCTTCCGGAATTGGCATTTCCACATGATGGCGCA 2913  
Qy 2933 CAGATCATCACCAAGACACCCGACCTGGGCTTGCCCACTCAATAACCACTCTTA 2992  
Db 2914 CAGATCATCACCAAGACACCCGAACTGGGCCCTGCCCACTCAACAACAACCACTCTTA 2973  
Qy 2993 CAAGCAATCTCCAGTCTTCAAGGGGCGCAGCAACGACAACCACTACTTGGCTACAG 3052  
Db 2974 CAACAATTTCCAGCAATCA---GGAGCTCGAACGACAATCACTACTTGGCTACAG 3030  
Qy 3053 CACCCCTGGGGTATTGATTTCACAGATTCCACTGCCACTTTTCAACCAAGTACTG 3112  
Db 3031 CACCCCTGGGGTATTGATTTCACAGATTCCACTGCCACTTTTCAACCAAGTACTG 3090  
Qy 3113 GCAGCACTCATCAACAACAATTGGGATTCGCGCCCAAGAGACTCAACTTCAAACTCTT 3172  
Db 3091 GCAAGACTCATCAACAACAAGTTCGGGATTCGACCCCAAGAGACTCAACTTCAAGCTCTT 3150  
Qy 3173 CAACATCCAAAGTCAAGAGTCAAGCAAGATGATGGCTCAACCAATGCTAATAACT 3232  
Db 3151 TAACTTCAAGTCAAGAGTCAAGCAAGATGACGGTACGACGATTTGCCAATAACT 3210  
Qy 3233 TACCAACAGGTTCAAGTCTTCTCGGACTCGAGTACCACTTCCGTACGCTCGGCTC 3292  
Db 3211 TACCAACAGGTTCAAGTCTTCTCGGACTCGAGTACCACTTCCGTACGCTCGGCTC 3270





Db 283 CTCCATTTGAAGCGGGAGGTTTGAACGCGCAGCCCGCATGCCGGGTTTACGAGATTG 342  
QY 357 TGATCAAGTGCCGAGCGACCTGAGCAGACCTGCCGGCATTTCTGACTCGTTGTGA 416  
Db 343 TGATTAAGTCCCAAGCAGCCTTGACGGGCACTGCCGGCATTTGACAGCTTTGTGA 402  
QY 417 GCTGGTGCCGAGAAAGAAATGAGCTGCCCCCGATTCTGACATGGATCTGAATCTGA 476  
Db 403 ACTGGGTGGCCGAGAAAGAAATGAGATTGCCGCCAGATTCTGACATGGATCTGAATCTGA 462  
QY 477 TTGAGCAGGCACCCCGTGACCGTGCCGAGAAAGCTGACGCGCAGCTTCTGTCCAATGGC 536  
Db 463 TTGAGCAGGCACCCCGTGACCGTGCCGAGAAAGCTGACGCGCAGCTTCTGACGGAATGGC 522  
QY 537 GCCCGGTGAGTAAAGCCCCCGAGGGCCCTCTTTCTTTGTTCAGTTGAGAAAGGGCGAGTCT 596  
Db 523 GCCGTGTGAGTAAAGCCCCCGAGGGCCCTTTTCTTTGTGCAATTTGAGAAAGGAGAGAGCT 582  
QY 597 ACTTCCACCTCCATATTTCTGTGTGAGAACCAAGGGGTCAAAATCCATGTGTGCGCGCT 656  
Db 583 ACTTCCACATGCACTGTCTCGTGAAGAACCAAGGGGTGAATCCATGTGTTTGGGACGTT 642  
QY 657 TCCTGAGTCAGATTAGGACAAAGCTGTGAGAACCATCTACCGCGGATCGAGCCGACCC 716  
Db 643 TCCTGAGTCAGATTGCGGAAAACTGATTCAGAGAAATTACCGCGGATCGAGCCGACTT 702  
QY 717 TGCCCAACTGTTCGCGGTGACCAAGACGCGTAAATGGCGCCGAGAGGGGGAACAAGTGG 776  
Db 703 TGCCAAACTGTTCGCGGTGACCAAGAACCAAGAAATGGCGCCGAGAGGGGGAACAAGTGG 762  
QY 777 TGAGCAGATGCTACATCCCAACTACTCTCTGCCCCAAGACTCAGCCCGAGCTGCAGTGG 836  
Db 763 TGATGAGTGTCTACATCCCAATTACTGTCTCCCAAAACCAGCTGAGCTCCAGTGG 822  
QY 837 CGTGACTTAACATGAGAGATATATAAGCGCTGTTGAACTGTGCGAGCGCAAAACGCG 896  
Db 823 CGTGACTTAATGGAACAGTATTTAAGCGCTGTTGAATCTCAAGAGCGTAAACGCT 882  
QY 897 TCGTGCGCAGCACTTGACCCAGCTGACGAGCAAGCAAGAGAGAGAGCAAGAAATCTGA 956  
Db 883 TGTGGCGCAGCATCTGACGACGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGATCAGA 942  
QY 957 ACCCAATCTGACGCGCTGTCTATCCGCTCAAAAACCTCCGCGCTACATGAGCTGG 1016  
Db 943 ATCCCAATCTGATGCGCGGTGATCAAGTCAAAAACCTTCAAGCCAGTACATGAGCTGG 1002  
QY 1017 TCGGTGCTGTGTGAGACCGGGCATCACTCCGAGAAAGCAAGTATCCAGAGAGAGAG 1076  
Db 1003 TCGGTGCTGTGTGAGCAAGGGGATTACTCGGAGAAAGCAAGTATCCAGAGAGAGAG 1062  
QY 1077 CCTGTACATCTCTTCAACGCGCTTCAACTCGCGGTCCAGATCAAGGCGCTGTGG 1136  
Db 1063 CCTCATCATCTCTTCAATGCGGCTTCAACTCGCGGTCCCAATCAAGGCTGCTTGG 1122  
QY 1137 ACAATGCGGCAAGATCATGCGCTGACCAAAATCCGCGCCGACTACCTGTAGGACCCG 1196  
Db 1123 ACAATGCGGAAAGATATAGCCTGACTAAACCGCCCGACTACCTGTGCGCAGC 1182  
QY 1197 CTCCGCGCGGACATTAAGAACCAACGCACTACCGCATCTGAGCTGAACGGCTAGC 1256  
Db 1183 AGCCCGTGAGAGACATTTCCAGCAATCGGATTTATAAATTTTGAACTAAACGGGTACG 1242  
QY 1257 AACCTGCTACGCGCGCTCGTCTTCTGTGCTGAGCTGGGCCAGAAAGGTTCCGGAAAGCGCA 1316  
Db 1243 ATCCCAATATGCGGCTTCCGCTTTCTGTGGATGGGCCACGAAGAAAGTTCCGCAAGAGGA 1302  
QY 1317 ACACCATCTGCTGTTTGGGCGCGGACCAAGGGCAAGACCAACATCGCGGAAGCCATCG 1376  
Db 1303 ACACCATCTGCTGTTTGGGCTTGAATACCGGGAAGACCAACATCGCGGAGGCCATAG 1362  
QY 1377 CCCAGCGCGTCCCTTCTACAGGCTGCGTCACTGGAACCAATGAGAACTTCCCTTCAATG 1436  
Db 1363 CCCACACTGTCCCTTCTACAGGCTGCGTAAACTGGAACCAATGAGAACTTCCCTTCAACG 1422

QY 1437 ATTGCGTCGACAAAGATGTGATCTGTGGGAGAGGGCAAGATGACGGCCAAAGTCTGTG 1496  
Db 1423 ACTGTGTGACAAAGATGTGATCTGTGGGAGAGGGGGAAGATGACCGCCAAAGTCTGTG 1482  
QY 1497 AGTCCGCCAAGGCCATTCTCGCGGAGACCAAGGTGCGCGTGAACCAAAAGTCAAGTCTG 1556  
Db 1483 AGTCCGCCAAGGCCATTCTCGAGGAAGCAAGGTGCGCGTGAACCAAAATGCAAGTCTCT 1542  
QY 1557 CCGCCAGATGACCCCAACCCCGTGTATGCTCACTTCAACCAACATGTGCGCGTGA 1616  
Db 1543 CCGCCAGATGACCCCAACCCCGTGTATGCTCACTTCAACCAACATGTGCGCGTGA 1602  
QY 1617 TTGACGGGAACAGCACCACTTCGAGCACCAAGAGCCGTTGAGAGACCGGATGTTCAAT 1676  
Db 1603 TTGACGGGAACCAACGACCTTCGAAACACAGAGCCGTTGAGAGACCGGATGTTCAAT 1662  
QY 1677 TTGAACCTACCCCGCTGTGAGCATGACTTTGGCAAGTGAACAAGCAGGAAGTCAAG 1736  
Db 1663 TTGAACCTACCCCGCTGTGATCACTGATCTTGGGAAGTGAACAAGCAGGAAGTCAAG 1722  
QY 1737 AGTCTTCCGCTGGGCGAGATCACTGACCGAGGTGCGCATGAGTTCTACGTCAGAA 1796  
Db 1723 ACTTTTCCGCTGGGCAAGGATCACTGTTGAGGTGAGCATGAATTTCTACGTCAGAA 1782  
QY 1797 AGGTGGAACCAACAAAGACCCGCCCGATGACGCGGATGAATAAGCAGGCCCAAGCGG 1856  
Db 1783 AGGTGGAACCAAGAAAGACCCGCCCGATGACGCGGATGAATAAGCAGGCCCAAGCGG 1842  
QY 1857 CTTGCCCTCAGTCGCGGATCACTGACGTCAGACGCGGAGAGAGCTCCGCTGACTTTG 1916  
Db 1843 TGCGGAGATCACTGCGAGCATGACGTCAGACGCGGA---AGCTTGATCAACTACG 1899  
QY 1917 CCGACAGTACCAAAACAATTTCTCGTCAACGCGGCACTGCTCAGATGCTGTTCCCT 1976  
Db 1900 CAGACAGTACCAAAACAATTTCTCGTCACTGCGGCACTGATGCTGTTCCCT 1959  
QY 1977 GCAAGCATGCGAGAGATGAATCAGAAATTTCAACATTTGCTTCAACGAGGAGAGAG 2036  
Db 1960 GCAAGCATGCGAGAGATGAATCAGAAATTTCAACATTTGCTTCACTCAGGAGAGAGAG 2019  
QY 2037 ACTGTTCAAGTGTCTTCCCGCGGTGTCAGAAATCTCAACCGT---CGTCAGAAAGAG 2093  
Db 2020 ACTGTTCAAGTGTCTTCCCGCGGTGTCAGAAATCTCAACCGT---CGTCAGAAAGAG 2076  
QY 2094 CGTATCGGAACTCTGTGCAATTCATCATCTGCTGGGCGGCTCCGAGATGCTGTGCT 2153  
Db 2077 CGTATCAGAACTGTGCTACATTCATCATATCATGGAAGAGTGCCAGA---CGCTGCA 2133  
QY 2154 CGGCTGCGATCTGTGCAACGTGACCTGATGACTGTGTTCTGAGCAATAATGACTT 2213  
Db 2134 CTGCTGCGATCTGTGCAATGTGATTTGATGACTGCACTTTGAAACAATAATGATTT 2193  
QY 2214 AAACAGGTATGCTGCGGATGTTATCTTCCAGATTGCTGAGAGCAAACTCTGTGAG 2273  
Db 2194 AAATCAGGTATGCTGCGGATGTTATCTTCCAGATTGCTGAGAGCACTCTCTGAA 2253  
QY 2274 GGCATTGCGAGTGTGGAATTGAACCTGAGGCCCGGAAGCCCAAGCCCAAGCAGCA 2333  
Db 2254 GGAATGAAGACAGTGTGAGGCTCAAACTGCGCCACCAACCAAGCCCGCAGAGCGG 2313  
QY 2334 AAGCAGAGAGAGCGCGGCTGTGCTTCTGCTCAAGTAACCTCGAGACCTTGAAC 2393  
Db 2314 CATTAAGAGAGAGAGGCTGTGCTTCTGCTCAAGTAACCTCGAGACCTTGAAC 2373  
QY 2394 GGACTGCACAAGGGGAGCGCTCAACGCGGAGAGAGCGCGGCTCGAGACG-ACAA 2452  
Db 2374 GGACTGCACAAGGGAGAGCGCTCAACGAGGAGAGAGCGCGGCTCGAGACGTAACA 2433  
QY 2453 GGCCTACGACGAGAGCTCAAGCGGGTGAACATCCGTAACCTGCGTATACCAAGCCGA 2512  
Db 2434 AGCTTACGACGAGAGCTGAGACGCGGAGACCAACCGTACTCAAGTACCAACGAGCGCA 2493



QY	2513	CGCCGAGTTTCAGGAGCGCTCTGCAGAAGATACGTCCTTTTGGGGGCAACCTCGGCGGAC	2572
Db	2494	CGCGGAGTTTCAGGAGCGCTTTAAAGAATACGTCCTTTTGGGGGCAACCTCGACGAGC	2553
QY	2573	AGTCTTCCAGGCCAAGAGCGGGTCTCGAACCTCTCGGTCTGTTGAGGAAGCGCTAA	2632
Db	2554	AGTCTTCCAGGCGGAAAAAGAGGGTCTTGAACCTCTGGGCGTGTGAGGAACCTGTAA	2613
QY	2633	GACGGCTCCTGAAAAGAACGTCGCGTAGAGCAGTCCGCCAAGAGCCAGACTCCTCTC	2692
Db	2614	GACGGCTCCGGGAAAAAGAGCGCGGTAGAGCACTCTCTGTGAGCCAGACTCCTCTC	2673
QY	2693	GGGCATCGGCAAGACAGCGCCAGACCCCGCTAAAAAGAGCTCAATTTTGTGACGCTG	2752
Db	2674	GGGAACCGGAAAGCGCGGCCAGACCTCGAAGAAAAAGATTGAATTTTGTCAGACTGG	2733
QY	2753	CGACTCAGAGTCAGTCCCCCGATCCACAACCTCTCGGAGAACCTCCAGCAACCCCGCTGC	2812
Db	2734	AGACGCAAGACTCAGTACCTGACCTCCAGCCTCTCGGACAGCTACCAAGAGCCCTCTGG	2793
QY	2813	TGTGGGACCTACTACAATGGCTTCAGGCGCGGTGGCGCACCAATGGCAGACAATAACGAGG	2872
Db	2794	TCTGGAACTAATACGATGGCTACAGGCAGTGGCGCACCAATGGCAGACAATAACGAGG	2853
QY	2873	CGCCGACGAGTGGGTAATGCTTCAGGAAATTGGCATTGGCATTCACATGGCTGGCGA	2932
Db	2854	CGCCGACGAGTGGGTAATTCCTCCGAAATTGGCATTGGCATTCACATGGATGGCGA	2913
QY	2933	CAGAGTCATCACCAACAGCACCCGACCTGGGCTTGCCACCTTACAATAACCACTCTA	2992
Db	2914	CAGAGTCATCACCAACAGCACCCGACCTGGGCTTGCCACCTTACAACAACCACTCTA	2973
QY	2993	CAAGCAATATCTCCAGTGCCTCAACGGGGGCCAGCAACGACAACCACTACTTGGGCTACG	3052
Db	2974	CAACCAATTTTCCAGCCAATCA--GAGCCTCGAACGACAATCACTACTTGGGCTACG	3030
QY	3053	CACCCCTGGGGGTATTTTGATTTCAACAGATTCCAACCTGCCACTTTTACCAACGCTG	3112
Db	3031	CACCCCTGGGGGTATTTTGACTTCAACAGATTCCACTGCTCACTTTTACCAACGCTG	3090
QY	3113	GCAGCACTCATCAACAACAATTGGGGATTTCGGCCCAAGAGACTCAACTTCAAACTCTT	3172
Db	3091	GCAAGACTCATCAACAACAATGGGGATTTCGACCCAAGAGACTCAACTTCAAGCTCTT	3150
QY	3173	CAACATCCAAAGTCAAGAGGTCAAGCAATGATGAGCTCAAAACATCGCTAATAA	3232
Db	3151	TAAATTCAAAGTCAAAAGAGGTACGCAAGATGACGCTACGACGATTGCCAATAA	3210
QY	3233	TACCAAGCAGGTTCAAGTCTTTCGGACTCGGAGTACCAAGCTTCCGTAAGTCTCGGCTC	3292
Db	3211	TACCAAGCAGGTTCAAGTCTTTCGGACTCGGAGTACCAAGCTTCCGTAAGTCTCGGCTC	3270
QY	3293	TGCGCAACGAGGCTGCTCTCCCTCGGTTCCCGCGGACGCTGTTCATGATTCCGCAATACG	3352
Db	3271	GCGGCATCAAGGATGCTCTCCGCGGTTCCAGCAGACGCTTTCATGCTGTCACAGATGG	3330
QY	3353	CTAAGTGAAGCTCAACAATGGCAGCCAGCCGCTGGAGCTTCATCTTTTACTGCTGGA	3412
Db	3331	ATACTCAACCTGAACAAGGAGTCAAGCAGTAGGACGCTTTCATTTTACTGCTGGA	3390
QY	3413	ATAATTCCTTCTCAGATGCTGAACGGGCAACAATTACTTCACTACACTTGA	3472
Db	3391	GTAATTCCTTCTCAGATGCTGCTACCGAAACAATTACTTCACTACACTTGA	3450
QY	3473	GGAAGTGCCTTTCACAGCAGCTACGCGCACAGCCAGAGCTTGGACCGGCTGATGAATCC	3532
Db	3451	GGAAGTGCCTTTCACAGCAGCTACGCTACAGCCAGAGCTTGGACCGGCTGATGAATCC	3510
QY	3533	TCTCATCGAACCAATACCTGTATTACTGAACAGAACTCAAAATCAGTCCGGAAGTCCCA	3592
Db	3511	TCTCATCGAACCAATACCTGTATTACTGAACAGAAACAACATCCAAAGTGAACCAAC	3570
QY	3593	AAACAAGAACTTGTCTTTAGCCGTTGGGCTCCAGCTGGCAATGTCTGTACGCCAAAA	3652

Db	3571	GCAGTCAAGGCTTCAGTTTCTCAGGCCGAGCGAGTGACATTCCGGACCACTTAGGAA	3630
Qy	3653	CTGGCTACCTGGACCCCTGTTATCGCAGACGCGCTTCTAAACAACAAACAGACAACAA	3712
Db	3631	CTGGCTTCTGACCCCTGTACCGCCAGCAGCAGATCAAAAGACATCTCGGATTAACA	3690
Qy	3713	CAACAGCAATTTTACCTGGACTGCTGCTTCAAATATACCTCAATGGGGTGAATCCAT	3772
Db	3691	CAACAGTGAATACTCTGTGACTGAGCTACCAAGTACCACCTCAATGGCAGAGACTCTCT	3750
Qy	3773	CATCAACCCCTGGCACTGCTATGGCCCTCACACAAGAAGCAGCAAGCAAGTTCTTCCAT	3832
Db	3751	GGTGAATC-----CGGCCATGGCAAGCCACAGAAGCATGAAGAAAAGTTTTTCCCTCA	3804
Qy	3833	GAGCGGTGCATGATTTTGGAAAAGAGAGCGCCGGAAGCTTCAAACACTGTGATTTGACAA	3892
Db	3805	GAGCGGGTCTCATCTTTGGGAAGCAAGGCTCAGAAAAACMAATGTGAACATTGAAAA	3864
Qy	3893	TGTCATGATTACAGACGAAGAGAAATTAAGCCACTAACCCCTGTGGCCACCGAAAAGATT	3952
Db	3865	GGTCATGATTACAGACGAAGAGAAATCGGAACAACCAATCCCGTGGCTACGAGCAGTA	3924
Qy	3953	TGGGACCGTGGCACTCAATTTCCAGAGCAGACACAGACCCTGCGACCGGAGATGTGCA	4012
Db	3925	TGGTCTGTATCTACCAACCTCCAGAGAGGCAACAGACAGCTACCGCAGATGTCAA	3984
Qy	4013	TGCTATGGGAGCATTAACCTGGCATGTGTGGCAAGATAGAGCGTACTCTGAGGGTCC	4072
Db	3985	CACACAAAGGCGTTCTTCCAGGCATGTCTGGCAGACAGAGATGTACTCTTCAAGGGGCC	4044
Qy	4073	CATTGGGCGCAAAATTTCTCACACAGATGAGACTTTCAACCCGTCTCCTTTATGGCCGG	4132
Db	4045	CATCTGGGCAAGATTCCACACACGAGCAGACATTTTCAACCCCTCTCCCTCATGGGTGG	4104
Qy	4133	CTTTGACTCAAGAACCCGCTCTCAGATCTCATCAAAAACACGCGCTGTCTTCTGCGAA	4192
Db	4105	ATTGGAATTAAACACCTCTCTCCACAGATTTCTATCAAGAACACCCCGTACTGCGAA	4164
Qy	4193	TCCTCCGCGAGTTTTCAGCTACAAAGTTTGTCTTATTCATCAACCAATTACTCCACAGG	4252
Db	4165	TCCTTCGACCACTTCAGTGCAGCAAAAGTTTGTCTTCTCATCACACAGTATCCACGGG	4224
Qy	4253	ACA-AGTGAAGTGAATAATGAATGGAGCTGCAGAAAGAAAACAGCAAGCGCTGGAATC	4311
Db	4225	ACAAGGTGACCGGTGAGATCGAGTGGAGCTGCAGAAAGAAAACAGCAAGCGCTGGAATC	4284
Qy	4312	CCGAAGTGCAATACATCCCAATTATGCAAAATCTGCCAAGCTTGATTTTACTGTGACA	4371
Db	4285	CCGAATTCAGTACACTTCCAACTACACAACAAGTCTGTTAATCGTGACCTTACCGTGATA	4344
Qy	4372	ACAATGACCTTTATACTGAGCCTCGCCCCATTTGGCAACCCGTTACCTTACCCTCTCTGT	4431
Db	4345	CTAATGGCGTGTATTCAGAGCCTCGCCCCATTTGGCAACAGATTACTGACTCTGTAATCTGT	4404
Qy	4432	AATTACGTGTTAATCAATAAACCGGTTGATTCGTTTCAAGTGAACCTTTGGTCTCTGTC	4491
Db	4405	AATTGCTTGTAAATCAATAAACCGTTTAATTCGTTTCAAGTGAACCTTTGGTCTCTGTC	4464
Qy	4492	TTCTTATCTTATCGGTTTACCATGTTATAGCTTACACATTAAGTCTTGGTTCGCTTGC	4551
Db	4465	TTTCTTCTTATCTAGTTTCCATGCTACGTAGTAAGTATGATGCGGGTTAATCATTA	4524
Qy	4552	CGATAAAGACTTACGTCATCGGGTTACCCCTAGATGAGTGTGCCCACTCCCTCTCTG	4611
Db	4525	ACTACAAGGA-----ACCCTAGATGAGTGTGGCCACTCCCTCTCTG	4568
Qy	4612	CGCGCTCGCTGCTCGGTGGGGCCCTGCGGACCAAAAGGTCCGACAGACGGCAGAGCTCTGCT	4671
Db	4569	CGCGCTCGCTGCTCACTGAGGCGGGCGACCAAAAGGTCTGCCGACGCGCGGCTTTGCG	4628
Qy	4672	CTGCCGGCCCCACCGAGCGAGCGGACGCGCAGAGAGGAGTGGGCAA	4718

Db 4629 CCGGCGGCTCAGTGAGCGAGCGCGCAGAGAGGAGTGGCCAA 4675

RESULT 9

US-10-240-198-1

; Sequence 1, Application US/10240198  
; Publication No. US20030100115A1

; GENERAL INFORMATION:

; APPLICANT: BTG International Ltd

; APPLICANT: BEARD DR, PETER

; APPLICANT: RAJ DR, KENNETH

; TITLE OF INVENTION: CYTOTOXIC AGENTS

; FILE REFERENCE: 142184WO

; CURRENT APPLICATION NUMBER: US/10/240,198

; PRIOR FILING DATE: 2002-09-30

; PRIOR APPLICATION NUMBER: 0009887.1

; NUMBER OF SEQ ID NOS: 6

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 1

; LENGTH: 4675

; TYPE: DNA

; ORGANISM: adeno-associated virus 2

US-10-240-198-1

Query Match 64.8%; Score 3055.8; DB 15; Length 4675;

Best Local Similarity 79.9%; Pred. No. 0;

Matches 3779; Conservative 0; Mismatches 887; Indels 61; Gaps 13;

QY 1 TTGCCCACTCCCTCTCTGCGCGCTCGCTCGCTGCGGCTGCGGACCAAGTCCGC 60

Db 1 TTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGCGCGGCGACCAAGTCCGC 60

QY 61 AGACGCGAGAGCTCTGCTGCGCGCGCCCAACGAGCGAGCGCGCAGAGAGGAGTG 120

Db 61 CGACGCGCGGCTTTGCGCGCGCGCTCAAGTGAAGCGAGCGCGCAGAGAGGAGTG 120

QY 121 GGCACATCCATCACTAGGGGTAAATCCGGAAGCGCCCTCCACCGCTGCCGCTCAAGCTGA 180

Db 121 GCCAATCCATCACTAGGGGT-----TCTGAGAGGGGTGAGTCTGTA 163

QY 181 CGTAAATTACGTATAGGG--GAGTGTCTGTATTAGTGTCAAGTGTCTTTGC 237

Db 164 CGTAAATTACGTATAGGGGTAGGAGGTCTGTATTAGAGTCAAGTGTG-TTTTGC 222

QY 238 GACATTTTGCAGACCAACGTCGCGCAATTTAGGGTATATATAGCGGAGTGAAGCAGAGAT 297

Db 223 GACATTTTGCAGACCAACGTCGCGCAATTTAGGGTATTTAAGCCGAGTGAAGCAGAGAT 282

QY 298 CTCATTTTG-ACCGGAATTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGATCG 356

Db 283 CTCATTTTGAAGCGGAGGTTGAACGCGCAGAGAGAGAGAGAGAGAGAGATTTTACGAGATTG 342

QY 357 TGATCAAGTGTCCGATCG 416

Db 343 TGATTAAGTGTCCGATTTGTA 402

QY 417 GCTGGTGGCCGATCGA 476

Db 403 ACTGGTGGCCGATTTGA 462

QY 477 TTGAG 536

Db 463 TTGAG 522

QY 537 GCGGCTGAGTAAAGTCT 596

Db 523 GCGGCTGAGTAAAGTCT 582

QY 597 ACTTCACTCACTATATTTCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 656

Db 583 ACTTCACTCACTATATTTCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 642

QY 657 TCCTGAGTCAGATTAGGACAAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 716

Db 643 TCCTGAGTCAGATTAGGACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 702

QY 717 TGCCCAACTGTTGCGCGGTGACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 776

Db 703 TGCCCAACTGTTGCGCGGTGACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 762

QY 777 TGAGAGAGTGTACATCCCACTACCTCTGCCCCAAGACTAGCCCGAGCTGAGTGGG 836

Db 763 TGAGAGAGTGTACATCCCACTACCTCTGCCCCAAGACTAGCCCGAGCTGAGTGGG 822

QY 837 CGTGAAGTAACTGAGAGAGATATTAAGCGGCTGTTGAACCTGCGGAGCGCAAGCGG 896

Db 823 CGTGAAGTAACTGAGAGAGATATTAAGCGGCTGTTGAACCTGCGGAGCGTAAACGCT 882

QY 897 TCGTGGCGAG 956

Db 883 TCGTGGCGAG 942

QY 957 ACCCAATTCTGACGCGCTGTATCCGCTCAAAAACCTCCGCGCTACATGAGAGCTGG 1016

Db 943 ATCCCAATTCTGATGCGCGGTGATCAGATCAAAAACCTCAGCCAGTACATGAGAGCTGG 1002

QY 1017 TCGGGTGGCTGTGAGAGCGGCGATCACTCCGAGAGAGAGAGAGAGAGAGAGAGAG 1076

Db 1003 TCGGGTGGCTGTGAG 1062

QY 1077 CTTGTATCATCTCTTCAAGCGGCTTCCAACTGCGGCTCCAGATCAAGCGGCTTGG 1136

Db 1063 CTTGTATCATCTCTTCAAGCGGCTTCCAACTGCGGCTCCAACTCAAGCGGCTTGG 1122

QY 1137 ACAATGCGGCAAGATCATGCGGCTGACCAAAATCCGCGCGAGTCACTGTAGCGCCG 1196

Db 1123 ACAATGCGGCAAGATCATGCGGCTGACCAAAATCCGCGCGAGTCACTGTAGCGCCG 1182

QY 1197 CTCGCGCGGAGACATTAACCAACGAGATCACTGAGAGAGAGAGAGAGAGAGAGAG 1256

Db 1183 AGCCCGTGAAGACATTTCCAGCAATCGAGTTTATAAATTTTGAAGTAAACGGGTACG 1242

QY 1257 AACCTGCTACGCGGCTCGCTTTCTGCGCTGGGCCAGAAAAGTTCCGGAAGCGCA 1316

Db 1243 ATCCCAATATATGCGGCTTCCGCTTTCTGTGAGTGGGCCAGAAAAGTTCCGGAAGCGA 1302

QY 1317 ACACCATCTGCTGTTGGGCGGCGCAACAGCGGCAAGCAATCGCGGAAGCCATCG 1376

Db 1303 ACACCATCTGCTGTTGGGCGGCGCAACAGCGGCAAGCAATCGCGGAAGCCATAG 1362

QY 1377 CCCAGCGGCTGCTTCAAGGCTGCTGCACTGAGAGAGAGAGAGAGAGAGAGAGAG 1436

Db 1363 CCCAGCGGCTGCTTCAAGGCTGCTGCACTGAGAGAGAGAGAGAGAGAGAGAGAG 1422

QY 1437 ATTGCTGCAAGAGTGTATCTGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1496

Db 1423 ATTGCTGCAAGAGTGTATCTGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1482

QY 1497 AGTCCGCCAAGGAGGATTTCCGCGCGAGCAAGAGTGGCGTGAACCAAAAGTGAAGTCT 1556

Db 1483 AGTCCGCCAAGGAGGATTTCCGCGCGAGCAAGAGTGGCGTGAACCAAAAGTGAAGTCT 1542

QY 1557 CCGCCAGATGAG 1616

Db 1543 CCGCCAGATGAG 1602

QY 1617 TTGACGGGAACAGACCACTTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1676

Db 1603 TTGACGGGAACAGACCACTTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1662

QY 1677 TTGAATCACTCCGCGCTGTGAGAGATGACTTTGGCAAGTGAACAAGCAGAGAGTCAAG 1736

Db 1663 TTGAATCACTCCGCGCTGTGAGAGATGACTTTGGCAAGTGAACAAGCAGAGAGTCAAG 1722

QY 1737 AGTTCTTCCGCTGGGCGAGAGATCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAG 1796

Db	1723	ACTTTTCCGGTGGCCAAAGATCACGTGGTTGAGGTGAGCATGAATTTACGTCAAAA	1782
QY	1797	AGGGTGGAGCCACAAAGAAGACCCGCCCCCGATGACCGCGGATAAAGCGAGCCCAAGCGG	1856
Db	1783	AGGGTGGAGCCCAAGAAAGACCCGCCCCCAGTGAACGCAGATATAAGTAGAGCCCAACGGG	1842
QY	1857	CCTGCCCCCTCAGTCCGGGATCCATCGACGTGACAGCGCGGAAGAGCTCCGGTGGACTTTG	1916
Db	1843	TGCGCGAGTCAGTTGCGCAGCCATCGACGTGACAGCGCGGA--AGCTTCGATCAACTACG	1899
QY	1917	CCGACAGGTACCAAAAATAATGTTCTCGTCACGCGGGCATGCTCAGATGCTGTTCCCT	1976
Db	1900	CAGACAGGTACCAAAAATAATGTTCTCGTCACGTGGGCATGAATCTGATGCTGTTCCCT	1959
QY	1977	GCAAGACATGCGAGAGATGAATTCAGAATTTCAACATTTGCTTCACGCACGGGACGAGAG	2036
Db	1960	GCAGACAATGCGAGAGATGAATTCAGAATTTCAACATTTGCTTCACGCACGAGAGAG	2019
QY	2037	ACTGTTCAAGATGCTTCCCCGGCGTGTCAGATCTCAACCGGT--CGTCAGAAAGAGA	2093
Db	2020	ACTGTTTAGAGTGTCTTCC---CGTGTCAAGATCTCAACCCGTTTCTGTCTCAAAAAG	2076
QY	2094	CGTATCGGAACTCTGTGCCATTCATCATCTGCTGGGGCGGGCTCCCGAGATTGCTTCT	2153
Db	2077	CGTATCAGAACTGTGCTACATTCATCATATCATATGGAAGAGTGCCAGA--CGCTTGCA	2133
QY	2154	CGGCTCGCATCTGGTCAACGTGGAACCTGATGACTGTGTTTCTGAGCAATAATGACTT	2213
Db	2134	CTGCTCGCATCTGGTCAATGTGGAATTTGGATGACTGCACTTTGAACAATAATGATTT	2193
QY	2214	AAACCAGGTATGCTGCGCGATGTTATCTTCCAGATTTGCTCGAGACAACCTCTCGAG	2273
Db	2194	AAATCAGGTATGCTGCGCGATGTTATCTTCCAGATTTGCTCGAGACACTCTCTCGAA	2253
QY	2274	GGCATTCGCGAGTGTGGAATTGAAACCTGAGACCCCGAAGCCCAAGCCCAACGCA	2333
Db	2254	GGAATTAAGACAGTGTGGAAGCTCAAACTGGCCACCAACCAAGCCCGCAGAGCGG	2313
QY	2334	AAGCAGGACGACGCGCGGGTCTGTGCTTCTGTGCTACAGTACCTCGAACCTTCAAC	2393
Db	2314	CATAAGGACGACAGCAGGGGTCTGTGCTTCTGTGCTACAGTACCTCGAACCTTCAAC	2373
QY	2394	GGAATTCGACAAAGGGGAGCCCGTCAACGCGCGGACCGCAGCGCCCTCGAGCAG--ACA	2452
Db	2374	GGAATTCGACAAAGGGGAGAGCCCGTCAACGAGCAGACGCGCGCGCCCTCGAGCAGTACA	2433
QY	2453	GGCTTACGACCGACAGCTCAAAAGCGGGTGACAATCCGTACTGCGGTATAACCAACGCGGA	2512
Db	2434	AGCTTACGACCGGACAGCTCGACAGCGGAGACAACCCGTACTCAAGTACAACCAACGCGGA	2493
QY	2513	CGCCGAGTTTCAGAGCGCTCTGCAAGAAGATACGTTTGGGGGCAACCTCGGCGAGC	2572
Db	2494	CGCGAGTTTCAGAGCGCGCTTAAAGAATACGTTTGGGGGCAACCTCGGCGAGC	2553
QY	2573	AGTCTTCCAGGCAAGAAAGCGGTTCTCGAACCCTCGGTCTGTTGAGGAAGCGCTAA	2632
Db	2554	AGTCTTCCAGGCAAGAAAGCGGTTCTTGAACCTCTGGGCTGTGAGGAACCTGTTAA	2613
QY	2633	GACGCGTCTCGGAAAGAAACGTCGCGGTAGAGCAGTCCGCCACAGAGCCAGACTCCTCTC	2692
Db	2614	GACGCGTCTCGGAAAGAAAGCGGTTAGAGCACTCTCTGTGAGAGCCAGACTCCTCTC	2673
QY	2693	GGGCATCGGCAAGACAGGCGCAGACCCGCTAAAAAGAGACTCAATTTTGGTCAGACTGG	2752
Db	2674	GCGAACCAGAAAGCGGCGCAGACCTGCAAGAAAAAGTTGAATTTTGGTCAGACTGG	2733
QY	2753	CGACTCAGAGTACGTCCCGATTCACAACCTCTCGGAGAACCTCCAGCAACCCCGCTGC	2812
Db	2734	AGACGAGACTCAGTACCTGACCCCAACCTCTCGGACAGCAACAGCAGCCCTCTGG	2793
QY	2813	TGTGGACCTACTACAATGCTTCAGGCGGTGCGCACCAATGGCAGACAATAACGAGG	2872

Db	2794	TCTGGGA	CTAATA	TGCGAT	GTGGCT	ACAGC	AGTGGCG	CAACCA	ATTGGC	AGACCA	ATTAA	CGAGGG	2853	
QY	2873	CGCCG	ACG	GAGT	GGTAA	TGCT	CTCAG	GAAT	TGGCA	TTGG	CA	TTGCCA	ATGCGTGGCGG	2932
Db	2854	CGCCG	ACG	GAGT	GGTAA	TTCCT	CCG	AAAT	TGGCA	TTGCCA	TTCC	ACAT	GATGGCGG	2913
QY	2993	CAGAGT	CAT	CACCA	CCAGCA	CCCGCA	CCCTGG	CGCTTG	CCCACT	TAC	CAAT	AA	CACTCT	2992
Db	2914	CAGAGT	CAT	CACCA	CCAGCA	CCCGCA	CCCTGG	CGCTTG	CCCACT	TAC	CAAT	AA	CACTCT	2973
QY	2993	CAAGCA	AATCT	CAAGT	CTTCA	ACG	GGG	CCAGCA	CGA	CAAC	CACT	ACTT	CGCT	3052
Db	2974	CAACAA	TAAT	TCCAG	CAAT	CA--	GGAG	CTCG	AA	CGA	CAAT	CACT	ACTT	3030
QY	3053	CACCCC	CTGG	GGGTAT	TTT	TGAT	TTCA	ACAG	AT	TCC	ACT	TTT	TCC	3112
Db	3031	CACCCC	CTGG	GGGTAT	TTT	TGAT	TTCA	ACAG	AT	TCC	ACT	TTT	TCC	3090
QY	3113	GCAGC	GA	CTCAT	CA	ACA	CAAT	TGG	GGAT	TCC	CG	CCCA	AGAG	3172
Db	3091	GCAAA	GACT	CAT	CA	ACA	CACT	GGG	AT	TCC	AG	CCCA	AGAG	3150
QY	3173	CAACAT	CCA	AGTCA	AGAG	GGTCA	CGA	CGA	ATGA	TGG	CGTCA	CAAC	CACTG	3232
Db	3151	TAA	CATT	CAAGTCA	AGAG	GGTCA	CGA	CGA	ATGA	TGG	CGTCA	CAAC	CACTG	3210
QY	3233	TACCAG	CA	CGGTTCA	AGTCTT	CTCG	GA	CTCG	GA	ATCA	CACT	CCGT	CGGCTC	3292
Db	3211	TACCAG	CA	CGGTTCA	AGTCTT	CTCG	GA	CTCG	GA	ATCA	CACT	CCGT	CGGCTC	3270
QY	3293	TGCGC	ACC	AGGCTG	CTCC	CTCC	GTTC	CCG	CGG	CGA	CGT	GTTC	ATGAT	3352
Db	3271	GGCGC	AT	CAAG	ATGCTCC	CGCG	GTTC	CC	AG	CA	AGT	CTTC	ATGAT	3330
QY	3353	CTACCT	G	ACGCTCA	ACA	TGG	CA	CGC	CA	AGC	CGT	GGG	ACGTT	3412
Db	3331	ATACT	CA	CCCTG	AA	CA	CGG	AGT	CA	GGC	AGT	AG	ACGCT	3390
QY	3413	ATA	TTT	CCCTT	CTC	AG	ATG	CTG	AA	AC	GGG	CA	CACTT	3472
Db	3391	GTA	CTT	CCCTT	CTC	AG	ATG	CTG	AA	AC	GGG	CA	CACTT	3450
QY	3473	GGA	AGT	GCCTT	CCA	CA	GA	CTAC	CG	CA	AGC	CGA	CGCT	3532
Db	3451	GGA	CGT	TCCTT	CCA	CA	GA	CTAC	CG	CA	AGC	CGA	CGCT	3510
QY	3533	TCT	CAT	CGA	CCA	TA	CTGT	ATTAC	CTG	AA	CA	GA	CTCA	3592
Db	3511	TCT	CAT	CGA	CCA	TA	CTGT	ATTAC	CTG	AA	CA	GA	CTCA	3570
QY	3593	AAA	CA	AGAC	TTGCT	TTAG	CCG	TG	GTCT	CA	GTG	GC	ATGT	3652
Db	3571	GCA	GT	CA	AGGCTT	CA	GT	TTCT	CA	GGCC	GA	CGTGA	CACTT	3630
QY	3653	CTG	GCT	ACCT	GG	ACCT	GT	TAT	TCG	CA	AGC	CGC	TTT	3712
Db	3631	CTG	GCT	ACCT	GG	ACCT	GT	TAT	TCG	CA	AGC	CGC	TTT	3690
QY	3713	CA	AC	GA	CAAT	TTT	TAC	CTG	GA	CTG	GTCTT	CA	AAAT	3772
Db	3691	CA	AC	GA	CAAT	TTT	TAC	CTG	GA	CTG	GTCTT	CA	AAAT	3750
QY	3773	CAT	CA	ACC	CTG	CA	CTGT	ATG	GCCT	CA	CA	AA	GA	3832
Db	3751	G	GT	GAATC	-----	CGG	CCAT	G	CA	AGC	CA	AGC	ATGA	3804
QY	3833	GAG	CG	GT	GCAT	GA	TTT	TG	AAAA	AG	AG	CG	CGG	3892
Db	3805	GAG	CG	GT	GCAT	GA	TTT	TG	AAAA	AG	AG	CG	CGG	3864
QY	3893	TG	T	CAT	GAT	TAC	AG	CA	GA	GA	AA	TTAA	AG	3952
Db	3865	G	G	T	CAT	GAT	TAC	AG	CA	GA	GA	AA	TTAA	



QY 3953 TGGACCGTGGAGTCAATTTCCAGAGCAGACAGACCCTGCGACCGAGATGTGCA 4012  
 Db 3925 TGGTCTGTATCTACCAACCTCCAGAGAGGCAACAGACAGCTACCGCAGATGTCAA 3984  
 QY 4013 TGCTATGGAGCATTAACCTGGCATGTGTGGCAAGATAGAGACGTGTACTGACGGTCC 4072  
 Db 3985 CACACAGGCGTTCTTCCAGGATGTGTGGAGGACAGAGATGTACTTCCAGGGGCC 4044  
 QY 4073 CATTGGGCGCAAAATCTCTCACACAGATGACATTTACCCCTCTCTCTATGGCGG 4132  
 Db 4045 CATCTGGCAAGATTCACACACAGGAGCATTTTCAACCCCTCTCTCTATGGGTGG 4104  
 QY 4133 CTTCGACTCAAGAACCCGCTCTCAGATCTCATCAAAAACAGCGCTGTCTCTGCGAA 4192  
 Db 4105 ATTGAGCTTAACACACCTCTCTCAGATCTCATCAAGACACCCGCTTACCTGCGAA 4164  
 QY 4193 TCCTCCGCGAGTTTTCAGCTACAAAGTTTCTTATTCATCAACCAATACTCCACAGG 4252  
 Db 4165 TCCTTCGACCACTTTCAGTGGCGCAAGTTTCTTCTTCTCATCACAGTACTCCACGGG 4224  
 QY 4253 ACA-AGTGTGTGAAATTTGAATGGAGCTGCAAGAAAGAAAACAGCAAGCGCTGAATC 4311  
 Db 4225 ACACGCTCAGCTGGAGATCGAGTGGAGCTGCAAGAAAGAAAACAGCAAGCGCTGAATC 4284  
 QY 4312 CCGAAGTGCAGTACACATCCCAATTATGCAAAATCTGCCAAGTTGATTTTACTGTGACA 4371  
 Db 4285 CCGAATTCAGTACACTTCCAACTACAAAGTCTGTTAATCTGTGACTTACCGTGATA 4344  
 QY 4372 ACAATGACTTATTAATGAGCTCGCCCAATGGCAACCGTTACCTTACCCGCTGCT 4431  
 Db 4345 CTAATGCGGTGTATTCAGAGCTCGCCCAATGGCAACAGATACCTGATCTGTAATCTGT 4404  
 QY 4432 AATTACGTGTATCAATAAACCGGTGATGTTTCACTGTAAGTTGCTCTCTGCTCC 4491  
 Db 4405 AATTGCTGTATCAATAAACCGTTTAACTTCACTGTAAGTTGCTCTCTGCTCA 4464  
 QY 4492 TTCTTATCTTATCGGTATCAATGATGTTATGCTTACATTAATCTGCTGTGCTGCT 4551  
 Db 4465 TTCTTATCTTATCGGTATCAATGATGTTATGCTTACATTAATCTGCTGTGCTG 4524  
 QY 4552 CGATAAAGACTTACGTATCGGTTACCCCTAGTATGATGAGTGGCCACTCCCTCTCTG 4611  
 Db 4525 ACTACAAGG-----ACCCCTAGTATGAGTGGCCACTCCCTCTCTG 4568  
 QY 4612 CGGCTGCTGCTCGCTGCTGCGGCTGCGCAAAAGTCCGACAGCGCAGAGCTCTGCT 4671  
 Db 4569 CGGCTGCTGCTCGCTCACTGAGCGCGGCAAAAGTCCGACAGCGCAGAGCTCTGCT 4628  
 QY 4672 CTGCGGCGCCGACGAGCGGAGCGCGCAGAGAGGAGTGGGCA 4718  
 Db 4629 CGGCGGCGCTGAGTGAAGGAGCGGCGCAGAGAGGAGTGGGCA 4675

RESULT 10  
 US-10-205-942-5  
 : Sequence 5, Application US/10205942  
 : Publication No. US20030053990A1  
 : GENERAL INFORMATION:  
 : APPLICANT: University of No. US20030053990A1th Carolina-Chapel Hill  
 : APPLICANT: Rabinowitz, Joseph E.  
 : APPLICANT: Samulski, Richard J.  
 : APPLICANT: Xiao, Weidong  
 : TITLE OF INVENTION: VIRUS VECTORS AND METHOD OF MAKING AND ADMINISTERING THE SAME  
 : FILE REFERENCE: 5470-186  
 : CURRENT APPLICATION NUMBER: US/10/205,942  
 : NUMBER OF SEQ ID NOS: 59  
 : SOFTWARE: Patent version 3.1  
 : SEQ ID NO 5  
 : LENGTH: 8179  
 : TYPE: DNA  
 : ORGANISM: Adeno-associated virus

US-10-205-942-5  
 Query Match 62.4%; Score 2941.8; DB 15; Length 8179;  
 Best Local Similarity 81.2%; Pred. No. 0;  
 Matches 3517; Conservative 0; Mismatches 792; Indels 20; Gaps 8;  
 QY 205 GTCCGTATTAAGTGTACAGTGTGCTTTTGGACATTTTGGACACCAAGTGGCAT 264  
 Db 121 GTCCGTATTAAGTGTACAGTGTGCTTTTGGACATTTTGGACACCAAGTGGCAT 179  
 QY 265 TAGGTAATATGCGCGAGTGAAGCAGAGATCTCCATTTG-AACCGCAATTTGAAC 323  
 Db 180 CTGGTAATTAAGCCGAGTGAAGCAGAGGCTCTCCATTTGAAGCGGAGTTGAAC 239  
 QY 324 GAGCAGACCATGCGCGCTTCTACAGATCGTGTATCAAGTGGCGGACCGCTGACG 383  
 Db 240 GCGAGCGCCATGCGCGCTTCTACAGATCGTGTATCAAGTGGCGGACCGCTGACG 299  
 QY 384 AGCAGTCCCGGCAATTTCTGACTGTTGTGAGCTGGTGGCGGAGAAAGATGGAGC 443  
 Db 300 GGCATCTGCGCGCAATTTCTGAGCTTTGTGAATGGTGGCGGAGAAAGATGGAGT 359  
 QY 444 TGCCCCCGAATTTGACATGATGATCTGATTTGAGCAGGACCCCTGACCGTGGCG 503  
 Db 360 TGCGCGCAGATTTGACATGATGATCTGATTTGAGCAGGACCCCTGACCGTGGCG 419  
 QY 504 AGAGTGCAGCGCGCACTTCTGCTCAATGCGCGCGGTGAAGGCCCCGAGGCGCC 563  
 Db 420 AGAGTGCAGCGCGCACTTCTGAGCAATGCGCGCGGTGAAGGCCCCGAGGCGCC 479  
 QY 564 TCTTCTTTGTTCACTGAGAGGCGGAGTCTTCTTCTTCTTCTTCTTCTTCTTCT 623  
 Db 480 TTTTCTTTGTTCACTGAGAGGCGGAGTCTTCTTCTTCTTCTTCTTCTTCTTCT 539  
 QY 624 CCAGCGGCTCAATCCATGCTGCTGGCGCGCTTCTTCTTCTTCTTCTTCTTCTTCT 683  
 Db 540 CCAGCGGCTCAATCCATGCTGCTGGCGCGCTTCTTCTTCTTCTTCTTCTTCTTCT 599  
 QY 684 TGCAGACATCTACCGCGGATGAGCGGACCCCTGCGCAACTGCTTCCGCTGACCAAG 743  
 Db 600 TGCAGACATCTACCGCGGATGAGCGGACCCCTGCGCAACTGCTTCCGCTGACCAAG 659  
 QY 744 CGCGTATGCGCGCGGAGGCGGAAACAAGTGTGAGCAGTGTACATCCCAATTAAC 803  
 Db 660 CGCGTATGCGCGCGGAGGCGGAAACAAGTGTGAGTGTACATCCCAATTAAC 719  
 QY 804 TCTGCGCAAGACTGAGCGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 863  
 Db 720 TCTGCGCAAGACTGAGCGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 779  
 QY 864 GCGCTGTTGAACCTGCGCGGAGGCGGAAACAAGTGTGAGTGTGAGTGTGAGTGT 923  
 Db 780 GCGCTGTTGAACCTGCGCGGAGGCGGAAACAAGTGTGAGTGTGAGTGTGAGTGT 839  
 QY 924 GCCAGACCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 983  
 Db 840 GCCAGACCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 899  
 QY 984 GGTCAAAAACCTCGCGCGGTACATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 1043  
 Db 900 GATCAAAAACCTCGCGCGGTACATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 959  
 QY 1044 CTTCCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1103  
 Db 960 CTTCCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1019  
 QY 1104 CCAACTGCGGTCCAGATCAAGCGCTCTGAGCAATGCGGCAAGATGAGCGCTGA 1163  
 Db 1020 CCAACTGCGGTCCAGATCAAGCGCTCTGAGCAATGCGGCAAGATGAGCGCTGA 1079  
 QY 1164 CCAATCCGCGCGGAGTGTGAGGCGCGCTGCGCGGAGCAATTAACCAACC 1223  
 Db 1080 CTAACCGCGCGGAGTGTGAGGCGCGCTGCGCGGAGCAATTAACCAACC 1139

QY	1224	GCATCTACCGCATCCTGGAGCTGAACGGGCTACGAACCTGCTTACGGCGGCTCCGCTTTTC	1283
Db	1140	GGATTATATAAATTTTGGAACTAAACGGGTACGATCCCAATATGCGGCTTCGCTTTTC	1199
QY	1284	TCCGCTGGGCCAGAAAAGTTCCGGGAAGCGCAACACACTCTGCTGTTTGGCGCGCA	1343
Db	1200	TGGGATGGGCCACGAAAAAGTTCCGCAAGAGAAACACCACTCTGGCTGTTTGGGCTGCAA	1259
QY	1344	CCACGGGCAAGACCAACACTCCGGGAAGCCATCGCCCACGCCCTTCTACGGCTGCG	1403
Db	1260	CTACCGGGAAAGACCAACACTCCGGAGGCCATAGCCCACTGTGCCCTTCTACGGGTGCG	1319
QY	1404	TCAACTGGACCAATGAGACTTTCCCTTCAATGATTGCGTCAACAAGATGATCTGAT	1463
Db	1320	TAAACTGGACCAATGAGACTTTCCCTTCAACGACTGTGTCAACAAGATGATCTGAT	1379
QY	1464	GGGAGGAGGCAAGATGACGGCCAAAGTCTGTGAGTCCGCCAAGGCCATTCTCGCGCA	1523
Db	1380	GGGAGGAGGCAAGATGACGGCCAAAGTCTGTGAGTCCGCCAAGCCATTCTCGAGGAA	1439
QY	1524	GCAAGTGCCTGTGACCAAAAGTGCAAGTCTCCGCCAGATGACCCCAACCCCGTGA	1583
Db	1440	GCAAGTGCCTGTGACCAAAAGTGCAAGTCTCCGCCAGATGACCCCAACCCCGTGA	1499
QY	1584	TCGTCACTTCCACACCAACACTGTGCGCCGTGATTGACGGGAAACGACCACTTCCGAC	1643
Db	1500	TCGTCACTTCCACACCAACACTGTGCGCCGTGATTGACGGGAACTCAACGACCTTCCGAC	1559
QY	1644	ACCAGCAGCCGTGTGACGACCGGATGTTCAATTTGAATCAACCCCGTCTGAGCATG	1703
Db	1560	ACCAGCAGCCGTGTGACGACCGGATGTTCAATTTGAATCAACCCCGTCTGAGCATG	1619
QY	1704	ACTTTGGCAAGGTGACAAAGCAGAAAGTCAAAAGTCTTCCGCTGGCGCAGATCAG	1763
Db	1620	ACTTTGGGAAGTCAACCAAGCAGAAAGTCAAAAGCTTTTCCGCTGGGCAAAAGATCAG	1679
QY	1764	TGACCGAGGTGCGCATGAGTTCTACGTCAGAAAGGTGAGCCAACAAAAGACCCGCC	1823
Db	1680	TGTTTGAAGGTGAGCATGATTTCTACGTCAAAAGGTGAGCCAAAGAAAGACCCGCC	1739
QY	1824	CCGATGACCGCGATTAAGCCGACCCCAAGCGGCTGCCCTTCACTGCGGATCCATCA	1883
Db	1740	CCAGTGACCGAGATATAAGTAGACCCCAACCGGGTCCGAGTCACTTGCAGGCCATCA	1799
QY	1884	CGTCAGACCGCGAAGGAGCTCCGGTGAATTGCCCCGACAGGTACCAAAACAAATGTTTC	1943
Db	1800	CGTCAGACCGCGA--AGCTTCGATCAACTACGACAGAGTACCAAAACAAATGTTTC	1856
QY	1944	GTCAACGCGGCGCTTCAAGTGTGTTTCCCTGCAAGACATGCGAAGAAATGAATCAGA	2003
Db	1857	GTCAACGCGGCGCTTCAAGTGTGTTTCCCTGCAAGACATGCGAAGAAATGAATCAGA	1916
QY	2004	ATTTCACATTTGCTTCAAGCAGCGGAGCAGAGACCTGTTCAAGTGTCTCCCGGCGTGT	2063
Db	1917	ATTTCACATTTGCTTCAAGCAGCGGAGCAGAGACCTGTTCAAGTGTCTCTCCCGGCGTGT	1973
QY	2064	CAGAATCTCAACCGGT--CGTCAGAAAGAGGACGTATCGGAACTCTGTGCCATTCAATC	2120
Db	1974	CAGAATCTCAACCGGTCTGTCTGTCTCAAAAGGCGTATCAGAACTGTGTCTACATTCAATC	2033
QY	2121	ATCTGCTGGGGCGGCTCCCGAGATGTGCTGCTCGGCTGCGATCTGTCAACGTGACC	2180
Db	2034	ATATCATGGGAAAGGTGCCAGA--CGCTTGCACTGCTGCGATCTGTCAATGTGATTT	2090
QY	2181	TGGATGACTGTGTTTCTGAGCAATAATGACTTAAACCAAGTATGCTGCCGATGTTAT	2240
Db	2091	TGGATGACTGCACTTTTGAACATAAATGATTTAAATCAGGATATGCTGCCGATGTTAT	2150
QY	2241	CTTCCAGATTGGCTCGAGGACAACTCTCTGAGGGCATTCGCGAGTGTGGACTTGAAA	2300
Db	2151	CTTCCAGATTGGCTCGAGGACAACTCTCTCTGAGGAAATAGACAGTGTGGAAAGCTCAAA	2210

QY	2301	CCTGAGACCCCGAAGCCCAAGCCCAACCCAGCAAAAGCAGGACGACCGCGCGGTCTGCTG	2360
Db	2211	CCTGGCCACCAACCAAAAGCCCGCAGAGCGGCATAAGACGACAGACGAGGGTCTGTG	2270
QY	2361	CTTCCTGGCTACAAAGTACCTCGGACCCCTTCAACGGACTCGACAAGGGGAGCCGTCAC	2420
Db	2271	CTTCCTGGGTACAAAGTACCTCGGACCCCTTCAACGGACTCGACAAGGGAGAGCCGTCAAC	2330
QY	2421	GCGGCGGACGACGCGGCCCTCGAGCAGACGAAGGCTTACGACCAGCAGCTCAAGCGGGT	2480
Db	2331	GAGGACAGACGCGCGGCCCTCGAGCAGACGAACAAGCCTACGACCGCAGCTGACAGCGGA	2390
QY	2481	GACAATCCGTACTCGCGGTATAACCAACGCCGACGCCGAGTTTCAGAGCGTCTGAAGAA	2540
Db	2391	GACAACCCGTACTCTCAAGTACAACACCGCCGACCGGAGTTTCAGAGCGCCTTAAGAA	2450
QY	2541	GATACGCTTTTGGGGCAACCTCGGGCGAGCAGTCTTCCAGGCCAAGACGGGTCTC	2600
Db	2451	GATACGCTTTTGGGGCAACCTCGGACGAGCAGTCTTCCAGGCCAAGACGGGTCTT	2510
QY	2601	GAACCTCTCGGTCTGGTGAAGGAGCGCTAAGACGGCTCCTGGAAGAAAGTCCGGTA	2660
Db	2511	GAACCTCTGGGCTGGTGAAGAACCTGTTAAGCGGCTCCGGAAAAAGAGCCGGTA	2570
QY	2661	GAGCAGTCGCCACAAGACCCAGACTCCTCGGGCATCGCAAGACAGCGCAGACCC	2720
Db	2571	GAGCAGTCTCTGTGAGGCAGACTCCTCCTCGGAAACGGAAAGCGGGCCAGACCT	2630
QY	2721	GCTAAAAGAGACTCAATTTGGTCAGACTGCGGACTCAGAGTCAGTCCCGATCCACAA	2780
Db	2631	GCAAGAAAAAGATTGATTTGGTCAGACTGAGACGACAGACTCAGTACCTGACCCCCAG	2690
QY	2781	CCTCTCGAGAACCTCCAGCAACCCCGCTGCTGTGGGAACTACTACAATGGCTTCAGGC	2840
Db	2691	CCTCTCGAGACGCCACGACGACGCCCTCTGTGCTGGGAACTAATACGATGGCTACAGGC	2750
QY	2841	GGTGGCGCAACCAATGGCAGACAATTAACGAAGCGCGCAGCGAGTGGGTATGCTTCAGGA	2900
Db	2751	AGTGGCGCAACCAATGGCAGACAATTAACGAAGCGCGCAGCGAGTGGGTATTTCTCCGGA	2810
QY	2901	AATTGGCATTGCGATTCCACATGCGTGGGCGCAGAGTCATCACCAACGACCCGCACC	2960
Db	2811	AATTGGCATTGCGATTCCACATGATGGGCGCAGAGTCATCACCAACGACCCGCACC	2870
QY	2961	TGGGCTTGCCACCTACAATAACCACTCTACAGCAATCTCCAGTCTTCAACGGGG	3020
Db	2871	TGGGCTTGCCACCTACAATAACCACTCTACAGCAATCTCCAGTCTTCAACGGGG	2927
QY	3021	GCCAGCAACGACAACCACTACTTCGGCTACAGCACCCCTGGGGTATTTGATTCAAC	3080
Db	2928	GCCTCGAAGCAATCACTACTTTGGCTACAGCACCCCTGGGGTATTTGACTTCAAC	2987
QY	3081	AGATTCCACTGCGACTTTTCAACCAAGTGACTGGGACCGACTCATCAACAACAATTGGGGA	3140
Db	2988	AGATTCCACTGCGACTTTTCAACCAAGTGACTGGGAAAGACTCATCAACAACAAGTGGGGA	3047
QY	3141	TTCCGGCCCAAGAGACTCAACTTCAAACTCTTCAACATCCAAGTCAAGGAGTCAAGCAG	3200
Db	3048	TTCCGACCCCAAGAGACTCAACTTCAAGCTCTTTAACAATTCAAGTCAAGAGGTCAAGCAG	3107
QY	3201	AATGATGGCGTCAACAACCAATCGCTAATAACCTTACACGACGCGTCAAGTCTTCTCGGAC	3260
Db	3108	AATGACGGTACGACGAGATGGCCAATAACCTTACACGACGCGTCAAGTCTTCACTGAC	3167
QY	3261	TCCGAGTACCAAGCTTCCGTACGTCTCGGCTCTGGGCAACGAGGCTGCTCCCTCCGTTT	3320
Db	3168	TCCGAGTACCAAGCTTCCGTACGTCTCGGCTCGGGCAACGAGGCTGCTCCCGCGTTT	3227
QY	3321	CCGGCGGACGTGTTCAATGATTCGCGCAATACGGGCTACGCTCAACAATGCGACCA	3380
Db	3228	CCAGCGGACGTCTTCAATGATTCCTCAGTATGATATCTCACCCCTGAACAACGGAAGTCAA	3287
QY	3381	GCCGTGGAGCTTCACTCTTTACTGCTGCTGAATATTTCCCTTCTCAGATGCTGAAGACG	3440





Db	660	CCAGAAATGCGCGGAGCGGGGAACAAGGTGGTGAATGAGTGTACATCCCAATTACT	719
QY	804	TCCTGCCCAAGACTCAGCCCGAGCTGCAGTGGCGGTGACTAACATGAGAGATATATA	863
Db	720	TGCTCCCAAAACCAGCCTGAGCTCCAGTGGCGGTGACTAATATGGAACAATATTTAA	779
QY	864	GCGCCTGTTGAACCTGGCCGAGCGCAAAACGGCTCGTGGCCGACGACCTGACCCAGTCA	923
Db	780	GCGCCTGTTGAATCTCAGGAGCGTAACCGGTGGTGGCGCAGCATCTGACGCACTGT	839
QY	924	GCCAGACCCAGAGCAGAACAAAGAAATCTGAACCCCAATTCTGACGCGCTGTTCATCC	983
Db	840	CGCAGACGAGAGCAGAACAAAGAAATCAGAAATCCAAATCTGATGCGCGGTGATCA	899
QY	984	GGTCAAAAAACCTCCGCGCTACATGAGCTGTCGGGTGGCTGTGTGACCGGGCATCA	1043
Db	900	GATCAAAAACTTCAGCCAGTACATGAGCTGTCGGGTGGCTGTGTGACAAAGGGATTA	959
QY	1044	CCTCCGAGAGCAGTGAATCCAGAGAGCAGCCCTCGTACATCTCTTCAACGCGCTT	1103
Db	960	CCTCGAGAGAGTGAATCCAGAGAGCAGCCCTCATACATCTCTTCAATGCGGCT	1019
QY	1104	CCAATCGCGGTCCCAATCAAGCCGCTCTGACAAATGCGGCAAGATCATGGCGCTGA	1163
Db	1020	CCAATCGCGGTCCCAATCAAGGCTGCTTGGACAAATGCGGAAAGATTATGAGCTGA	1079
QY	1164	CCAAATCCGCGCCCGACTACCTGTGTAGCCCGCTCCGCGCGGACATTAACCAACC	1223
Db	1080	CTAAAAACGCCCCGACTACTGTGTGGCCAGCAGCAGCCGCTGAGAGACATTTCCAGCAATC	1139
QY	1224	GCATCTACCGCATCTGTAGCTGAACGCGCTACGAACCTGCTTACGCGGCTCGTCTTTC	1283
Db	1140	GGATTATTAATTTTGAATAAACGGGTACGATCCCAATATGCGGCTTCGCTTTTC	1199
QY	1284	TGCGCTGGGCCAGAAAAAGTTGCGGAAGCGCAACACATCTGCTGTTTGGCGCGCA	1343
Db	1200	TGGGATGGGCCAGAAAAAGTTGCGCAAGAGAACACCACTGGCTGTTGGCGCTGCAA	1259
QY	1344	CCACGGGCAAGACCAACATGCGGGAAGCCATCGCCCAAGCGGCTGCTTCAAGGCTGCG	1403
Db	1260	CTACCGGGAAGACCAACATCGCGGAGGCCATAGCCCACTGTGCCCTTCTACGGGTGCG	1319
QY	1404	TCAACTGGACCAATGAGAACTTTCCTTCAATGATTGCGTTCGAACAAGATGTGATCTGT	1463
Db	1320	TAAACTGGACCAATGAGAACTTTCCTTCAACGACTGTGTCCACAAGATGTGATCTGT	1379
QY	1464	GGAGGAGGGCAAGATGACGCGCAAGGTCGTGAGTCCGCCAAGGCCATTCTCGCGCA	1523
Db	1380	GGAGGAGGGGAAGATGACCGCCAAAGTCTGTGAGTCCGCCAAGCCATTCTCGAGAA	1439
QY	1524	GCAAGTGCAGTGACCAAAAGTGCAAGTCGTCCGCCAGATCGACCCCAACCCCGTGA	1583
Db	1440	GCAAGTGCAGTGACCAAAATGCAAGTCCTCGGCCAGATAGACCCGACTCCCGTGA	1499
QY	1584	TCGTCACTCCAACACCAACATGTCGCCCGCTGATTGACGGGAACAGCACCACTTCGAGC	1643
Db	1500	TCGTCACTCCAACACCAACATGTCGCCCGCTGATTGACGGGAACCAACGACCTTCGAC	1559
QY	1644	ACCAAGCAGCCGTTGCAAGACCGGATGTTCAATTTGAATTCACCCGCGCTGTGAGCATG	1703
Db	1560	ACCAAGCAGCCGTTGCAAGACCGGATGTTCAATTTGAATTCACCCGCGCTGTGATCATG	1619
QY	1704	ACTTTGGCAAGGTGACAAAGCAGGAAGTCAAAAGATTCTTCGCTGGGCGCAGGATCAG	1763
Db	1620	ACTTTGGGAAGTCAACAGCAGGAAGTCAAAAGACTTTTCCGCTGGGCAAGGATCAG	1679
QY	1764	TGACCGAGGTGCGCATGAGTTCTACGTCAAAAAGGTGAGACCAAAAAAGACCCGCC	1823
Db	1680	TGCTTGAAGTGAAGCATGAATCTACGTCAAAAAGGTGAGACCAAAAAAGACCCGCC	1739
QY	1824	CCGATGACCGGATAAAAAGCAGCCCAAGCGGCGTGGCCCTCAGTCGCGGATCATGGA	1883

Db	1740	CCAGTGACCGCAGATATAAGTGAGCCCAACCGGGTCCCGAGTCAGTTGCCGACCATCGA	1793
Qy	1884	CGTCAGACCGCGGAAGAGCTCCGGTGGACTTTGGCCGACAGGTACCMAAACAAATGTTCTC	1943
Db	1800	CGTCAGACCGCGGA--AGCTTCGATCAACTACGAGACAGGTACCMAAACAAATGTTCTC	1856
Qy	1944	GTCACGCGGGGCTGCTTCAGATGCTGTTCCCTGCAAGACATGCCAGAGAATGAATCAGA	2003
Db	1857	GTCACGTGGGCATGAATCTGATGCTGTTCCCTGCAAGCAATGCCAGAGAATGAATCAGA	1916
Qy	2004	ATTTCACATTTGCTTCACGACGAGACGAGACTGTTCAAGTGTCTTCCCGCGGT	2063
Db	1917	ATTCAATATCTGCTTCACTCACGAGACAGAAAGACTGTTAGAGTCTTTCC--CGTGT	1973
Qy	2064	CAGATCTCAACCGGT---CGTCAGAAAGAGACGTATCGAAACTCTGTGCCATTATC	2120
Db	1974	CAGATCTCAACCGGT---CGTCAGAAAGAGACGTATCGAAACTCTGTGTACATTCATC	2033
Qy	2121	ATCTGCTGGGCGGGCTCCCGAGATTGCTTGTCTGCGCTCGCATCTGTTCAAGTGAAC	2180
Db	2034	ATATCATGGGAAAGGTGCCAGA--CGCTTGCACTGCTCGCATCTGTTCAATGTGAT	2090
Qy	2181	TGGATGACTGTGTTTCTGAGCAATAATGACTTAAACAGGTATGGCTGCCGATGGTTAT	2240
Db	2091	TGGATGACTGCACTTTGAAACAATAATGATTAAATCAGGTATGGCTGCCGATGGTTAT	2150
Qy	2241	CTTCAGATTGGCTCGAGACAACCTCTCTGAGGGCATTCGCGAGTGTGGACTTGAAA	2300
Db	2151	CTTCAGATTGGCTCGAGACAACCTCTCTGAGGAATAAGACAGTGTGGAAGCTCAAA	2210
Qy	2301	CCTGGAGCCCGAAGCCCAAGCCCAACCAAGCAAAAGCAGACGACGCGGGGTCTGTG	2360
Db	2211	CCTGGCCACCAACCAACCAAGCCCGCAGAGCGGCATAAGACGACGAGGGGTCTGTG	2270
Qy	2361	CTTCCTGGCTACAAGTACTCGGACCCTTCAACGCACTCGACAAGGGGAGCCCGTCAAC	2420
Db	2271	CTTCCTGGGTACAAGTACTCGGACCCTTCAACGCACTCGACAAGGAGCCCGTCAAC	2330
Qy	2421	GCGGCGGACGCGCGCCCTCGAGCAGCACAAGGCTTACGACCAAGCTCAAGCGGGT	2480
Db	2331	GAGGCAAGCGCGCGCCCTCGAGCAGCACAAGGCTTACGACCAAGCTCAAGCGGGT	2390
Qy	2481	GACAATCCGTACTGCGGTAATAACCAAGCCGACGCGGAGTTTCAAGAGCGTCTGCAAGA	2540
Db	2391	GACAATCCGTACTGCGGTAATAACCAAGCCGACGCGGAGTTTCAAGAGCGGCTTCAAGG	2450
Qy	2541	GATACGCTTTTGGGGCAACCTCGGGCAGCAGTCTTCCAGGCCAAGAGCGGGTCTC	2600
Db	2451	GACACATCGTTTGGGGCAACCTCGGCAGAGCAGTCTTCCAGGCCAAGAGCGGGTCTT	2510
Qy	2601	GAACTCTCGGTCTGTTGAGGAAGGCGCTAAGACGGCTCTGGAAGAAACGTCCGGTA	2660
Db	2511	GAACTCTTGTGTGTTGAGCAAGCGGTGAGACGGTCTTGGAAGAAAGAGACCGTTG	2570
Qy	2661	GAGCAGTCGCCACAAGACCAAGACTCTCTCGGGCATTCGGCAAGACAGCGCACGCC	2720
Db	2571	ATTGAATCCCCCAGCAGCCCCGACTCTCCAAGGTATCGGCAAAAAAGCAGCAGCCG	2630
Qy	2721	GCTAAAAAGAGACTCAATTTTGGTCAGACTGGCGACTCAGAGTCAGTCCCCGATCCACA	2780
Db	2631	GCTAAAAAGAGCTCGTTT-----CGAAGCAGAACTGGAGCAGGCGCAGCGACCC	2681
Qy	2781	CCTCTCGAGAACCTCCAGCAACCCCGCTGCTGTGGACCTACTCAATGGCTTCAGGC	2840
Db	2682	CCTGAGGATCAACTTCGGAGCCATGTCTGATG-----ACAGTGAGATGCTGACGA	2735
Qy	2841	GGTGGCGCACCAATGGCAGACAATAACGAAGCGCCGACGAGTGGGTAATGCTCAGGA	2900
Db	2736	GCTGGCGGAGCTGCAGTCGAGGGCGGACACAGGTGCCGATGAGTGGGTAATGCTCGGGT	2795
Qy	2901	AATTGGCATTTGGATTCCATGGCTGGGCGCAGAGTTCATCACCAAGCACCAGCCGCAAC	2960
Db	2796	GATTGGCATTTGGATTCCATGGCTGGGCGGCGCAGCTTCACCAAGCACCAGCACCAGAAC	2855





OY	324	GAGCAGCAGCCATGCGGGCTTCTACGAGATCGTGATCAAGGTGCCGAGCCTTGACG	383
Db	4882	GCGCAGCCGCCATGCGGGTTTACGAGATGTGATTAAAGTCCCGACGACCTTGACG	4823
OY	384	AGCACCCTGCGGGCATTTCTGACTCGTTGTGAGCTGGGTGCGGAGAAGATGGAGC	443
Db	4822	GGCATCTGCCCGCATTTCTGACAGCTTTGTGAACCTGGGTGCGGAGAAGATGGAGT	4763
OY	444	TGCCCCCGGATTTCTGACATGAACTGAATCTGATTGAGCAGGTAACCCCTGACCGTGCCG	503
Db	4762	TGCCGCCAGATTCTGACATGAACTGAATCTGATTGAGCAGGTAACCCCTGACCGTGCCG	4703
OY	504	AGAACTGCAGCGCGCATTTCTCTGCTCCAATGGCGCCGCGTGAATAAGCCCCGGAGGCC	563
Db	4702	AGAACTGCAGCGCGCATTTCTCTGCTCCAATGGCGCCGCGTGAATAAGCCCCGGAGGCC	4643
OY	564	TCTTCTTTGTTCAGTTGAGAAAGGCGAGTCTTACTTCCACCTTCATATTCTGTGAGAA	623
Db	4642	TTTCTTTGTCAATTGAGAAAGGAGAGAGACTACTTCCACATGCACTGTCTGTGAAA	4583
OY	624	CCACGGGGTCAAAATCCATGTGCTGGCGCGCTTCTGAGTCAAGTAAGGACAAGCTTG	683
Db	4582	CCACGGGGTGAATCCATGTGTTTGGGACGTTCTTGAGTCAAGATTCGGAAAACTGA	4523
OY	684	TGCAGACCATCTACCGCGGATTCGAGCCGACCCCTGCCCACTGTTCCGGGTGACCAAGA	743
Db	4522	TTCAGAGAAATTACCGCGGATTCAGCCGACTTTGCCAACTGTTCCGGGTCACAAGA	4463
OY	744	CGCGTAATGGCGCGGAGGGGGGAAACAAGTGTGACGAGTGTACATCCCAACTACC	803
Db	4462	CCAGAAATGGCGCGGAGGGGAAACAAGTGTGATGAGTGTCTACATCCCAATTACT	4403
OY	804	TCCTGCCCAAGACTCAGCCCGAGCTGCAGTGGCGGTGACTAAATGAGAGATATATAA	863
Db	4402	TGCTCCCCAAAACCCAGCCTGAGCTCCAGTGGCGGTGACTAATATGAAACAGTATTAA	4343
OY	864	GCGCTGTTGAACCTGGCGGAGCGCAAAACGGCTCGTGGCGCAGCACTGACCCACGTCA	923
Db	4342	GCGCTGTTGAATCTCAGGAGGGTAACAAGTGTGTTGGCGAGCATCTGACGACGTGT	4283
OY	924	GCCAGACCCAGAGCAGAACAGAGAAATCTGAACCCCAATTCTGACGCGCTGTATCC	983
Db	4282	CGCAGACGCAAGAGCAGAACAGAGAAATCAGAATCCCAATTCTGATCGCGCGTGATCA	4223
OY	984	GGTCAAAAACCTCCGCGGCTACATGGAAGCTGTGGGTGGCTGTGGAACCGGGCATCA	1043
Db	4222	GATCAAAAACCTCAGCCAGGTACATGAGCTGTGGGTGCTGTGGAACAAGGGGATTA	4163
OY	1044	CCTCCGAGAGCAGTGAATCCAGAGGAGACCAAGCCTCGTACATCTCTTCAACGCGCTT	1103
Db	4162	CCTCGAGAGCAGTGAATCCAGAGGAGACCAAGCCTCATACATCTCTTCAATGCGCT	4103
OY	1104	CCAACCTCGCGTCCAGATCAAGGCGCTCTGACAATGCGCGCAAGATCATGCGCTGA	1163
Db	4102	CCAACCTCGCGTCCCAATCAAGGCTGCTTGACAATGCGGAAAGATTATGAGCCTGA	4043
OY	1164	CCAAATCCGCGCGCGACTACTCTGTAGGCCCCCGCTCCGCCCGGACAATTAAACCAACC	1223
Db	4042	CTAAAAACGCGCGCGACTACTCTGTGGGCCAGCAGCCCGTGAAGACAATTTCCAGCAATC	3983
OY	1224	GCATCTACCGCATCTTGAGCTGAACGGCTACGAACCTGCTTACGCGGCTCGTCTTTC	1283
Db	3982	GGATTTATAAATTTTGAACATAACGGGTACGATCCCAATAATGCGGCTTCCGCTTTC	3923
OY	1284	TGCGCTGGGCCAGAAAAAGTTCCGGAAAGCGCAACACCATGTGCTGTTTGGCGCGCA	1343
Db	3922	TGGGATGGGCCACGAAAAAGTTCCGGCAAGAGAACACCATGTGCTGTTTGGCGCTGCA	3863
OY	1344	CCACGGGCAAGACCAACATCGCGGAAGCCATGCCCCACGCGCTGCCCTTCTACGCGTGC	1403
Db	3862	CTACCGGGAAGACCAACATCGCGGAAGCCATGCCCCACACTGTGCCCTTCTACGCGGTGC	3803

QY	1404	TCAACTGACCAATGAGAACTTCCCTTCAATGATTCGGTCGACAAAGATGTGATCTGGT	1463
Db	3802	TAAACTGGACCAATGAGAACTTCCCTTCAACGACTGTGTGACAAAGATGTGATCTGGT	3743
QY	1464	GGAGAGGGGCAAGATGACGGCCAAAGTCGTGAGTCGCCCAAGGCCATTCTCGCGCA	1523
Db	3742	GGAGAGGGGGAAGATGACCGCCAAAGTCGTGAGTCGGCCAAAGCCATTCTCGAGAA	3683
QY	1524	GCAAGTGCCTGAGCAAAAGTGCAAGTCGTCCGCCAGATCGACCCACCCTCGTGA	1583
Db	3682	GCAAGTGCCTGAGCAAGAATGCAAGTCCTCGGCCAGATGACCCGACTCCGTGA	3623
QY	1584	TCGTCACCTCCAACACCAACATGTGCGCGGTGATTTGACGGGAAACAGACCACTTCGAGC	1643
Db	3622	TCGTCACCTCCAACACCAACATGTGCGCGGTGATTTGACGGGAACTCAACGACCTTCGAC	3563
QY	1644	ACCAAGCAGCCGTTCAGAGACCGGATGTTCAATTGAACTCACCCGCCGTCTGAGCATG	1703
Db	3562	ACCAAGCAGCCGTTCAGAGACCGGATGTTCAATTGAACTCACCCGCCGTCTGAGCATG	3503
QY	1704	ACTTTGGCAAGGTGACAAAGCAGGAAGTCAAAGAGTTCTTCCGCTGGGCGCAGGATCAG	1763
Db	3502	ACTTTGGCAAGGTGACCAAGCAGGAAGTCAAAGAGTTTTCGGTGGGCAAAAGATCAG	3443
QY	1764	TGACCGAGGTGGCGCATGAGTCTACGTCAAGAAAGGTGAGCCAAACAAAGACCCGCC	1823
Db	3442	TGAGTGGAGTGGAGCATGAAATTCAGTCAAAAGGGTGGAGCCAAAGAAAGACCCGCC	3383
QY	1824	CCGATGACGCGGATAAAGCGAGCCCAAGCGGGCTGCCCTCAGTCGCGGATCCATCGA	1883
Db	3382	CCAAGTACGCGAGATATAAGTGAAGCCCAACGGGTGGCGAGTCAGTTGGCAGCCATCGA	3323
QY	1884	CGTCAGACGCGGAAGAGCTCCGGTGGACTTTGCCGACAGGTACCAAAACAATGTTCTC	1943
Db	3322	CGTCAGACGCGGAAGAGCTCCGGTGGACTTTGCCGACAGGTACCAAAACAATGTTCTC	3266
QY	1944	GTCAAGCGGGCATGCTTCAGATGCTGTTCCCTGACAGACATGCGAGAGATGAATCAGA	2003
Db	3265	GTCAAGTGGGTATGAATCTGATGCTTTTCCCTGCGGCAATGCGAGAGATGAATCAGA	3206
QY	2004	ATTCAACATTGCTTCAACGCACGGACGAGAGACTGTTCAAGATGCTTCCCGCGCTGT	2063
Db	3205	ATGTGACATTGCTTCAACGCACGGGGTCAAGTGTGCGGAGTGTCCCTCCCTCCCT	3149
QY	2064	CAGAATCTCAAC-----CGTCTCAGAAAGAGAGCGTATCGAAACTCTGTGCCATTC	2117
Db	3148	CAGAATCTCAACCGCTGTCTGTCTCAGAAAGCGGACGTATCAGAAACTGTGTCCGATTC	3089
QY	2118	ATCATCTGCTGGGGCGGGTCCCGAGATTGCTTGTCTGGCTGCGATCTGTCAACGTGG	2177
Db	3088	ATCATCATCATGGGGAGGGGCGCCGAGGTGGCTGTCTGGCTGCGAATCGCCAAATGGG	3029
QY	2178	ACCTGATGACTGTGTTCTTGAGCAATAATGACTTAAACCAAGTATGCGTGCAGATGGT	2237
Db	3028	ACTTGATGACTGTGATGAGCAACAATAATGACTTAAACCAAGTATGACT--GACGCT	2972
QY	2238	TATCTTCCAGATTGGCTCGAGGACAACTCTCTGAGGGCATTCGCGAGTGTGGACTTG	2297
Db	2971	TACCTTCCAGATTGGCTAGAGGACAACTCTCTGAAGCGTTGAGAGTGTGGCGCTG	2912
QY	2298	AAACCTGAGCCCCCGAAGCCCAAGCCAAACAGCAAAAGCAGAGCAGCGCCGGGCTGTG	2357
Db	2911	CAACCTGAGCCCCCTAAACCCAAAGCAAAATCAACATCAGAGCAACGCTCGGGGTCTT	2852
QY	2358	GTGCTTCCCTGCTAGAGTACCTCGGACCTTCAACGGAATCGACAAAGGGGAGCCCGTC	2417
Db	2851	GTGCTTCCCGGTTACAAATACCTCGGACCCGCAACGGAATCGACAAAGGGGAAACCGTC	2792
QY	2418	AAAGCGGGCGGACGCGCGCTCGAGCAGCACAAGGCTTACGACGAGCTCAAAAGCG	2477
Db	2791	AAAGCAGCGGACGCGCGCGCTCGAGCAGCACAAGGCTTACGACGAGCTCAAGGCC	2732
QY	2478	GGTGAACAATCCGTACTGCGGTATTAACCAAGCCGACCGCGAGTTTCAAGAGCGGTTCGAA	2537



Db 2731 GGTGACAAACCCCTACCTCAAGTACAAACCAACCGGAGCTCCAGCAGCGGCTTCAG 2672  
Qy 2538 GAAGATACGCTTTTGGGGGCAACCTCGGGGCAAGCAGTCTTCCAGGCCAAGAAGCGGTT 2597  
Db 2671 GGGCACAACGCTTTGGGGGCAACCTCGGGGCAAGCAGTCTTCCAGGCCAAGAAGCGGTT 2612  
Qy 2598 CTGGAACCTCTGGGCTGTGTTGAGGAAGGCGCTAAGACGGCTCTGGAAGAAAGAGCTCCG 2657  
Db 2611 CTGGAACCTCTGGGCTGTGTTGAGGAAGGCGCTCTGGAAGAAAGAGAGACCG 2552  
Qy 2658 GTAGAGCAGTCGCCACAAGAGCCAGACTCTCTCGGGCATCGGCAAGAGCCAGCAG 2717  
Db 2551 TTGATGAAATCCCCCAGCAGCCGACTCTCCACGGGTATCGGCAAAAAGGCAAGCAG 2492  
Qy 2718 CCGGCTAAAAGAGACTCAATTTTGTGAGACTGGCGACTCAGAGTCAATCCCGATCCA 2777  
Db 2491 CCGGCTAAAAGAGACTCGTTT-----CGAAGACGAAACTGAGCAGAGCGACGGA 2441  
Qy 2778 CAACCTCTCGAGAAACCTCCAGCAACCCCGCTGCTGTGGGACCTACTACAATGGCTTCA 2837  
Db 2440 CCGGCTGAGGATCAACTTCCGAGGCCATGTCTGATG-----ACAGTGAATGCGTCA 2387  
Qy 2838 GGGGCTGGGCAACCAATGGCAGACAAATACGAAGGCGCCGAGCGAGTGGGTATGCTCA 2897  
Db 2386 GCAGCTGGGAGCTGCAGTCGAGGGCGGCAAGGTGCCGATGAGTGGGTATGCTCG 2327  
Qy 2898 GGAATTTGCAATGCGATTCCACATGCTGGGCGAGAGTCAATCAACCAAGCAGCCGC 2957  
Db 2326 GGTGATTTGCAATGCGATTCCACATGCTGGGCGAGAGTCAATCAACCAAGCAGCCGC 2267  
Qy 2958 ACCTGGGCTTGCCACCTACAATAACCACTCTACAAGCAAAATCTCAAGTGTTCACG 3017  
Db 2266 ACCTGGGCTTGCCACCTACAATAACCACTCTACAAGCAAAATCTCAAGTGTTCACG 2219  
Qy 3018 GGGGCGACGACGACCACTACTTCCGCTACAGCAGCCCGCTGGGGTATTTGATTTTC 3077  
Db 2218 GAGAGCTTCGACGCTCCAAACACTTCAACGGAATTTCTCACCCCGCTGGGGTATTTGATTTTC 2159  
Qy 3078 AACAGATTTCCACTGCGCACTTTTCAACGAGTGAAGTCAAGGAGTCAATCAACCAATTTG 3137  
Db 2158 AACCGCTTCCACTGCGCACTTTTCAACGAGTGAAGTCAAGGAGTCAATCAACCAATTTG 2099  
Qy 3138 GGATTTCCGCGCCCAAGAGACTCAACTTCAAACTCTTCAACATCCAAGTCAAGGAGTCAAG 3197  
Db 2098 GGCATGCAACCAAGGAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 2039  
Qy 3198 AACGAATGATGCGCTCAACCACTGCTAATACTTACCAGCAGGTTCAAGTCTTCTCG 3257  
Db 2038 ACCTGCAACGCGAGACAAACGAGTGAATACTTACCAGCAGGTTCAAGTCTTCTCG 1979  
Qy 3258 GACTCGAGTACCAAGCTTCCGTAAGTCTCGGCTCTGCGACAGAGGCTGCTCCCTCG 3317  
Db 1978 GACTCGAGTACCAAGCTTCCGTAAGTCTCGGCTCTGCGACAGAGGCTGCTCCCTCG 1919  
Qy 3318 TTCCCGGCGAGCCTGTTCAATGATTTCCGCAATACGGCTAC-----CTGACGCTCAAC 3368  
Db 1918 TTCCCGGCGAGCCTGTTCAATGATTTCCGCAATACGGCTAC-----CTGACGCTCAAC 1859  
Qy 3369 AATGCAAGCAGCGCTGAGCAGTCACTTTTACTGCTGGAATATTTCCCTTCTCG 3428  
Db 1858 ACTTGCAGCAACAGACTGACAGAAATGCTTCTACTGCTGAGTACTTTCTTCTCGCAG 1799  
Qy 3429 ATGCTGAGAAAGGCGCAACACTTACTTCACTACACTTTGAGGAAGTGCCTTCTCG 3488  
Db 1798 ATGCTGAGAAAGGCGCAACACTTACTTCACTACACTTTGAGGAAGTGCCTTCTCG 1739  
Qy 3489 AGCAGTACGCGCAGACGAGCGCTGAGCGGCTGATGAATCTCTCATCGACCAATAC 3548  
Db 1738 TCGATGTAAGCGCAGACGAGCGCTGAGCGGCTGATGAATCTCTCATCGACCAATAC 1679  
Qy 3549 CTGTATTAATCTGAA---CAGAACTCAAAATCAAGTCCGGAAGTCCCAAAACAGGACTTG 3605

Db 1678 CTGTGGGAGCTGCAATCGAACCAACCGGAACCAACCTGAAATGCCGGAGTGCACCAAC 1619  
Qy 3606 CTGTTAGCCGCTGGGCTCTCCAGCTGGCATGTCTGTTACAGCCCAAAAAGTGGTACCTGGA 3665  
Db 1618 AACTTAAACAGCTGCGGCTCACTTTTCCAACTTTAAAGAACTGGCTGCCGGG 1559  
Qy 3666 CCGGTTATCGGAGCAGCGCGTTTCTAAACCAAAAACAGCAACCAACAGCAATTTT 3725  
Db 1558 CTTCAATCAAGCAGAGGGGCTTCTCAAAAGACTGCCAATCAAACTAAGATCCCTGCC 1499  
Qy 3726 ACCTGCACTGGTCTTCAAAATTAAC-----CTCAATGGCGCTGAATCC 3770  
Db 1498 ACCGGCTGAGACAGTCTCATTAATACGAGACGACAGCACTTGAACGGAAGATGAGT 1439  
Qy 3771 ATCATCAACCTGGCACTGTATGCGCTCACACAAGACGACGAAGCAAGTCTTTCC 3830  
Db 1438 GCCCTGACCCCGGACCTCCATGCGCACGCGCTGACCTGGGACAGCAAGTTC--AGC 1382  
Qy 3831 ATGACCGGTGTCATGATTTTGGAAAAAGAGAGCGCGGAGCTTCAAACTGCAATGAGC 3890  
Db 1381 AACAGCCAGCTCATCTTTGGCGGCTTAAACAGAACGGAACAGCGCACCTACCCGGG 1322  
Qy 3891 AATGTCATGATTAACAGACGAAGGAATTAAGCCCACTAACCTGTGGCCACCGAAGA 3950  
Db 1321 ACTGTCATCTTCACTCTGAGAGAGAGCTGGCAGCCACCAACCGCATACGAGCATG 1262  
Qy 3951 TTTGGACCGTGGCAGTCAATTTCCAGAGCAGACGACCAACCTGCGACCGGAGATGTG 4010  
Db 1261 TGGGCAACCTACCTGCGGCTGACAGAGCAACAGCAACCTGCGACCGTGAACAGTGTG 1202  
Qy 4011 CATGCTATGGAGCATTACCTGGCATGTGTGGCAAGATAGAGCCTGTACCTGACGGGT 4070  
Db 1201 ACAGCCTTGGAGCCGTGCTGGAATGCTGTGGCAAAAGAGACATTTACTACAGGGT 1142  
Qy 4071 CCCATTGGGCGCAAAATCTCTCAACAGATGAGACACTTCAACCCGTCTCTTATGGGC 4130  
Db 1141 CCCATTGGGCGCAAAATCTCTCAACAGATGAGACACTTCAACCCGTCTCTTATGGGC 1082  
Qy 4131 GGGTTTGGACTCAAGAACCGGCTCTCAGATCTCTCAATCAAAAACAGCGCTTCTCGG 4190  
Db 1081 GGGTTTGGGCTGAACACCGGCTCTCAATTTTATCAAGAAACACCGGCTACCTGCG 1022  
Qy 4191 AATCTCCGCGGAGTTTCACTCAACAAGTTGCTTCAATTCATCAACCAATACCTCA 4250  
Db 1021 AATCTCCGCGGAGTTTCACTCAACAAGTTGCTTCAATTCATCAACCAATACCTCA 962  
Qy 4251 GGAAGTGAAGTGTGAATTTGAATGGAGCTGCAGAAAGAAACAGCAAGCGCTGAAT 4310  
Db 961 GGCAGGCTGCTGAGATTTGAATGGAGCTGCAGAAAGAAACAGCAAGCGCTGAAT 902  
Qy 4311 CCGGAAGTGCAGTACACATCAATTAATGCAAAATCTGCCAAGTGAATTTTACTGTGAC 4370  
Db 901 CCGGAAGTGCAGTACACATCAATTAATGCAAAATCTGCCAAGTGAATTTTACTGTGAC 842  
Qy 4371 AACATGACTTTATACAGCTCGCCCATTTGGCAGCCGTTACCTTACCCGCTCCGCTG 4430  
Db 841 GCGGCTGGAATATACATGAGCTAGGCTATCGGTACCCGCTACCTACCCACCACTG 782  
Qy 4431 TAATTAAGTGTATTAATTAACCGGTTGATTCGTTTCAATGAACTTTGCTGCTC--TGT 4489  
Db 781 TAATTAAGTGTATTAATTAACCGGTTTATTCGTTTCAATGAACTTTGCTGCTC--TGT 722  
Qy 4490 CTTTCTTATCTTATC--GGTTACCATGTTATAGCTTACAGA 4529  
Db 721 CTTTCTTATCTTATCTGTTTCCATGGCTACTGCTACATA 681

RESULT 13  
US-09-792-630-12  
; Sequence 12, Application US/09792630  
; Patent No. US20020168640A1  
; GENERAL INFORMATION:  
; APPLICANT: Li, Min

; APPLICANT: Dahiya, Bassil I.  
 ; TITLE OF INVENTION: BIOCHIPS COMPRISING NUCLEIC ACID/PROTEIN CONJUGATES  
 ; FILE REFERENCE: A-70295/RFT/RMS/RMK  
 ; CURRENT APPLICATION NUMBER: US/09/792,630  
 ; NUMBER OF SEQ ID NOS: 87  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 12  
 ; LENGTH: 1872  
 ; TYPE: DNA  
 ; ORGANISM: adeno-associated virus 1  
 ; US-09-792-630-12

Query Match 39.7%; Score 1872; DB: 11; Length 1872;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1872; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	335	ATGCCGGGCTTCTACGAGATCGTATCAAGTCCCGAGCCGACCTGACGACCACTGCCG	394
DB	1	ATGCCGGGCTTCTACGAGATCGTATCAAGTCCCGAGCCGACCTGACGACCACTGCCG	60
QY	395	GGCATTCTGACTCGTTTGTGAGCTGGGTGGCGAGAAAGAAATGGAGCTGCCCGGAT	454
DB	61	GGCATTCTGACTCGTTTGTGAGCTGGGTGGCGAGAAAGAAATGGAGCTGCCCGGAT	120
QY	455	TCTGACATGATCTGAATCTGATTGAGCAGGCACTTGAACCTGAGCCGTGGCGAGAAAGCTG	514
DB	121	TCTGACATGATCTGAATCTGATTGAGCAGGCACTTGAACCTGAGCCGTGGCGAGAAAGCTG	180
QY	515	CGGACTTCTGCTTCAATGAGCGCGCGGTGAGTAAGGCCCGAGGCCCTCTTTTGT	574
DB	181	CGGACTTCTGCTTCAATGAGCGCGCGGTGAGTAAGGCCCGAGGCCCTCTTTTGT	240
QY	575	CAGTTGAGAAAGGCGAGTCTCTACTTCCACTTCAATTTCTGTGAGACCAAGGGGTC	634
DB	241	CAGTTGAGAAAGGCGAGTCTCTACTTCCACTTCAATTTCTGTGAGACCAAGGGGTC	300
QY	635	AAATCCATGCTGCTGGCGCGCTTCTGAGTCAAGTAAAGGACAAAGCTGTGACAGCATC	694
DB	301	AAATCCATGCTGCTGGCGCGCTTCTGAGTCAAGTAAAGGACAAAGCTGTGACAGCATC	360
QY	695	TACCGCGGATCGAGCCGACCTGCGCAACTGCTTCCGGTGACCAAGCGCTAATGGC	754
DB	361	TACCGCGGATCGAGCCGACCTGCGCAACTGCTTCCGGTGACCAAGCGCTAATGGC	420
QY	755	GCCGAGGGGGGAAAGAGTGTGTGACGAGTGTACATCCCAACTACTCTGCCCCAG	814
DB	421	GCCGAGGGGGGAAAGAGTGTGTGACGAGTGTACATCCCAACTACTCTGCCCCAG	480
QY	815	ACTGAGCCGAGTCAAGTGTGGCGTGAATTAAGAGAGTATATAAGCGCTGTTG	874
DB	481	ACTGAGCCGAGTCAAGTGTGGCGTGAATTAAGAGAGTATATAAGCGCTGTTG	540
QY	875	AACCTGGCGGAGCGAAAGCGCTCTGGCGGACCTGACCCAGCTGACGACAGCCAG	934
DB	541	AACCTGGCGGAGCGAAAGCGCTCTGGCGGACCTGACCCAGCTGACGACAGCCAG	600
QY	935	GAGCAGAACAGAGAAATCTGAACCCCAATTCTGACGCGCTGTCAATCCGGTCAAAAAC	994
DB	601	GAGCAGAACAGAGAAATCTGAACCCCAATTCTGACGCGCTGTCAATCCGGTCAAAAAC	660
QY	995	TCCGCGGCTTACATGAGCTGTGCGGTGGTGTGGAACCGGGCATCACTCCGAGAG	1054
DB	661	TCCGCGGCTTACATGAGCTGTGCGGTGGTGTGGAACCGGGCATCACTCCGAGAG	720
QY	1055	CAGTGATCAGAGGACAGGCGCTGTACATCTCTTCAACGCGCTTCAACTGCGG	1114
DB	721	CAGTGATCAGAGGACAGGCGCTGTACATCTCTTCAACGCGCTTCAACTGCGG	780
QY	1115	TCCAGATCAAGCGCTCTGAGCAATGCGGAGATCATGCGCTGACCAATCCGCG	1174
DB	781	TCCAGATCAAGCGCTCTGAGCAATGCGGAGATCATGCGCTGACCAATCCGCG	840

QY	1175	CCGACTACCTGTAGCCCCCGCTCCGCCGCGACATTAACCAACCGCATCTACCG	1234
DB	841	CCGACTACCTGTAGCCCCCGCTCCGCCGCGACATTAACCAACCGCATCTACCG	900
QY	1235	ATCTGAGCTGAACGCGTACGAACCTGCTACGCGCGCTCTTCTCGGCTGGCC	1294
DB	901	ATCTGAGCTGAACGCGTACGAACCTGCTACGCGCGCTCTTCTCGGCTGGCC	960
QY	1295	CAGAAAGTTCCGGAAGCGCAACCATCTGCTGTTGGCCCGCCACCGAGGCAAG	1354
DB	961	CAGAAAGTTCCGGAAGCGCAACCATCTGCTGTTGGCCCGCCACCGAGGCAAG	1020
QY	1355	ACCAACATCCGGGAGCCATCCGCCACCGCGTCCCTTCTACGCGCTGCTCACTGACC	1414
DB	1021	ACCAACATCCGGGAGCCATCCGCCACCGCGTCCCTTCTACGCGCTGCTCACTGACC	1080
QY	1415	AATGAACTTTCCCTTCAATGATTGCGTCAACAAAGATGTGATCTGTGGAGGAGGC	1474
DB	1081	AATGAACTTTCCCTTCAATGATTGCGTCAACAAAGATGTGATCTGTGGAGGAGGC	1140
QY	1475	AAGATGACGGCCAAAGTGTGAGTCCGCCAAGGCCATTCTGGCGGAGCAAGTGGC	1534
DB	1141	AAGATGACGGCCAAAGTGTGAGTCCGCCAAGGCCATTCTGGCGGAGCAAGTGGC	1200
QY	1535	GTGACCAAAAGTGCAAGTGTGCGCCAGATGCAACCCCACTGATGTCACCTCC	1594
DB	1201	GTGACCAAAAGTGCAAGTGTGCGCCAGATGCAACCCCACTGATGTCACCTCC	1260
QY	1595	AACACCAATGTGCGCGGTGATGACGGGAACAGACCACTTGAAGCAACGAGCGC	1654
DB	1261	AACACCAATGTGCGCGGTGATGACGGGAACAGACCACTTGAAGCAACGAGCGC	1320
QY	1655	TTGACGACCGGATGTTCAAAATTGAATCAACCCCGCTGAGCATGACTTTGGCAAG	1714
DB	1321	TTGACGACCGGATGTTCAAAATTGAATCAACCCCGCTGAGCATGACTTTGGCAAG	1380
QY	1715	GTGACAAAGCAGGAAGTCAAAAGTCTTCCGCTGGCGGACGATCAGTGAACGAGTG	1774
DB	1381	GTGACAAAGCAGGAAGTCAAAAGTCTTCCGCTGGCGGACGATCAGTGAACGAGTG	1440
QY	1775	GCGCATGAGTCTACGTCAAGAAAGGTGAGCCAAACAAAGACCGCCCGATGACCG	1834
DB	1441	GCGCATGAGTCTACGTCAAGAAAGGTGAGCCAAACAAAGACCGCCCGATGACCG	1500
QY	1835	GATAAAGCAGACCCCAAGCGGCTGCCCCCTAGTCCGGAATCCATGACGTGACGCG	1894
DB	1501	GATAAAGCAGACCCCAAGCGGCTGCCCCCTAGTCCGGAATCCATGACGTGACGCG	1560
QY	1895	GAAAGCTCCGCTGACTTTGCCGACAGGTACCAAAACAAATGTTCTGTCAGCGGCG	1954
DB	1561	GAAAGCTCCGCTGACTTTGCCGACAGGTACCAAAACAAATGTTCTGTCAGCGGCG	1620
QY	1955	ATGCTTACATGCTGTTTCCCTGCAAGACATGCGAGAGAAATGAATCAGATTCAACAT	2014
DB	1621	ATGCTTACATGCTGTTTCCCTGCAAGACATGCGAGAGAAATGAATCAGATTCAACAT	1680
QY	2015	TGCTTACGCAAGGAGCAGAGACTGTTCAGAGTGTCTCCCGGCGTGTCAAGATCTCA	2074
DB	1681	TGCTTACGCAAGGAGCAGAGACTGTTCAGAGTGTCTCCCGGCGTGTCAAGATCTCA	1740
QY	2075	CCGCTGTCAGAAAGAGACGTATCGAAACTCTGTGCCATTATCATCTGCTGGGCGG	2134
DB	1741	CCGCTGTCAGAAAGAGACGTATCGAAACTCTGTGCCATTATCATCTGCTGGGCGG	1800
QY	2135	GCTCCGAGATTGCTTGTGCTGCGCTGCGATCTGTCACAGTGAACCTGATGACTGT	2194
DB	1801	GCTCCGAGATTGCTTGTGCTGCGCTGCGATCTGTCACAGTGAACCTGATGACTGT	1860
QY	2195	TCTGAGCAATAA 2206	
DB	1861	TCTGAGCAATAA 1872	







OY	2015	TGCTTCACGCA	CGGGA	CGAGAGACTGTT	CAGAGTGCCTT	CCCCCGCGTGT	CAGAATCTCA	2074
Db	1681	TGCTTCACGCA	CGGGA	CGAGAGACTGTT	CAGAGTGCCTT	CCCCCGCGTGT	CAGAATCTCA	1740
OY	2075	CCGGTCGTCA	GAAGA	GAGACGATCCG	AACTCTGTGC	CAATTCATCTG	CTGGGCGG	2134
Db	1741	CCGGTCGTCA	GAAGA	GAGACGATCCG	AACTCTGTGC	CAATTCATCTG	CTGGGCGG	1800
OY	2135	GCTCCCGAG	ATGCTGCTCG	GCCCTGCGAT	CTGTCAACG	TGACCTGATG	ACTGTGTT	2194
Db	1801	GCTCCCGAG	ATGCTGCTCG	GCCCTGCGAT	CTGTCAACG	TGACCTGATG	ACTGTGTT	1860
OY	2195	TCTGAGCA	ATAA	2206				
Db	1861	TCTGAGCA	ATAA	1872				

Search completed: July 20, 2003, 20:11:27  
Job time : 1032.56 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2003, 07:55:16 ; Search time 12440.8 Seconds  
(without alignments)  
11036.790 Million cell updates/sec

Title: US-09-807-802A-1

Perfect score: 4718

Sequence: 1 ttgccacatccctctctgcg.....cgagagagggagtgggca 4718

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 segs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:

1: gb\_ha:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_cm:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_cm:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_mus:\*  
34: em\_htg\_pln:\*  
35: em\_htg\_rod:\*  
36: em\_htg\_mam:\*  
37: em\_htg\_vrt:\*  
38: em\_sy:\*  
39: em\_htgo\_hum:\*  
40: em\_htgo\_mus:\*  
41: em\_htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	4718	100.0	4718	14	AF063497	AF063497 Adeno-ass
2	4258	90.3	4683	14	AF028704	AF028704 Adeno-ass
3	3211.6	68.1	4726	14	AVU48704	U48704 Adeno-assoc
4	3205	67.9	4722	14	AF028705	AF028705 Adeno-ass
5	3107.8	65.9	4679	6	AX282480	AX282480 Sequence
6	3107.8	65.9	4679	14	AF043303	AF043303 Adeno-ass
7	3107.8	65.9	8698	6	AX205072	AX205072 Sequence
8	3082.4	65.3	4680	6	AR028767	AR028767 Sequence
9	3082.4	65.3	4680	6	I62303	I62303 Sequence 1
10	3055.8	64.8	4675	6	AX135805	AX135805 Sequence
11	3055.8	64.8	4675	6	AX286292	AX286292 Sequence
12	3055.8	64.8	4675	14	AA2CG	J01901 Adeno-assoc
13	2993.4	63.4	7557	6	AX205073	AX205073 Sequence
14	2878.4	61.0	7327	12	AF369963	AF369963 Cloning v
15	2598.6	55.1	4767	14	AVU89790	U89790 Adeno-assoc
16	1691.2	35.8	4072	6	AX205074	AX205074 Sequence
17	1472.2	31.2	2116	14	AA2LEFT	J01902 adeno-assoc
18	1425.8	30.2	4910	6	AR027100	AR027100 Sequence
19	1425.8	30.2	4910	6	AR036062	AR036062 Sequence
20	1425.8	30.2	4910	6	AR160052	AR160052 Sequence
21	1425.8	30.2	4910	6	AR209764	AR209764 Sequence
22	1354.2	28.7	4642	14	AF085716	AF085716 Adeno-ass
23	1354.2	28.7	4652	6	AX256321	AX256321 Sequence
24	1352.6	28.7	4404	14	AAV18065	Y18065 adeno-assoc
25	1008	21.4	9258	12	AF383623	AF383623 Cloning v
26	877.2	18.6	4101	14	AF416726	AF416726 Goose par
27	861.6	18.3	5132	14	BDU22967	U22967 Barbarie du
28	860	18.2	5106	14	GPU25749	U25749 Goose parvo
29	852	18.1	4550	14	MDENSVP	X75093 Muscovy duc
30	795.8	16.9	3434	14	GPU34761	U34761 Goose parvo
31	734	15.6	2307	6	AX256327	AX256327 Sequence
32	729.8	15.5	2264	6	AX256328	AX256328 Sequence
33	729.8	15.5	2264	6	AX256329	AX256329 Sequence
34	690.6	14.6	1690	6	AX256335	AX256335 Sequence
35	690.6	14.6	1870	6	AX256331	AX256331 Sequence
36	616	13.1	2487	14	MDPVPS	Z68272 Muscovy duc
37	476	10.1	1115	6	AX256333	AX256333 Sequence
38	475	10.1	1292	6	AX256330	AX256330 Sequence
39	433.8	9.2	606	6	AX344105	AX344105 Sequence
40	399	8.5	653	14	XXAAV1	V01457 Messenger R
41	262.8	5.6	579	14	HUAAV01	M27085 Homo sapien
42	229.2	4.9	539	14	AF166104	AF166104 Goose par
43	229.2	4.9	539	14	AF166105	AF166105 Goose par
44	229.2	4.9	539	14	AF166106	AF166106 Goose par
45	227.6	4.8	539	14	AF166102	AF166102 Goose par

#### ALIGNMENTS

RESULT 1  
AF063497  
LOCUS AF063497 4718 bp DNA linear VRL 27-APR-1999  
DEFINITION Adeno-associated virus 1, complete genome.  
ACCESSION AF063497  
VERSION AF063497.1 GI:4689096  
KEYWORDS  
SOURCE adeno-associated virus 1.  
ORGANISM adeno-associated virus 1.  
REFERENCE 1 (bases 1 to 4718)  
AUTHORS Xiao, W., Chirmule, N., Berta, S.C., McCullough, B., Gao, G. and Wilson, J.M.  
TITLE Gene therapy vectors based on adeno-associated virus type 1  
JOURNAL J. Virol. 73 (5), 3994-4003 (1999)

MEDLINE 99214338  
PUBMED 10196295  
REFERENCE 2 (bases 1 to 4718)  
AUTHORS Xiao, W. and Wilson, J. M.  
TITLE Direct Submission  
JOURNAL Submitted (05-MAY-1998) IHGT, University of Pennsylvania, 3601 Spruce Street, Philadelphia, PA 19104, USA

FEATURES  
source Location/Qualifiers  
1. 4718  
/organism="adeno-associated virus 1"  
/db\_xref="taxon:85106"  
/note="AAV1"  
335. 2206  
/codon\_start=1  
/product="nonstructural protein"  
/protein\_id="AAD27758.1"  
/db\_xref="GI:4689098"

CDS  
1. 4718  
/translation="MPGFYEIVIKVPSDLDEHLPGISDSFVSWVAEKEMELPDSMD  
LNLIEQAPLTVAEKLQDFLQWRVSKAPEALFVQFEKESYFHLHLVETGVKS  
MVLGRFLSQIRDLVQTIYRGIEPTLPMFAVTKTRNGAGGNKVDECIYNYLIPK  
TOPELQWMTNMEEYISACLAEKRLVAQHLTHSOTOEONKLNPNSDAPVIRS  
KTSARIMELVGLVDRGITSSEKQIQEDQASISFNAASNSRQIKALDNAGKIMAL  
TKSAPDYLVGPAPADIKTNRIYRLNGLYEPAYAGSVFLGAKRFGKRTIMLFG  
PATGKTINIAEIAHAVPYGCNMTNENFPNDQVDMVWEEGKMTAKVESAKA  
ILGSKVRVDQCKSSAQIDPTPIVTSNTNMCVIDNSTFEHQQLQDRMFKFEL  
TRRLEHDFGKVTQEVKEFRWAQDHTVEVAHLEFVRKGGANKRPADADAKPKRA  
CPVADPSTSDAEGAPVDFADRYQNKSRHAGMLQMLFPCKTCERNQNFNICTHGT  
RDCSECFPGVSESQPVVRKRTYRKLCAIHLLGRAPEIACSACDLVNVLDLDCVSEQ"  
2223. 4433  
/codon\_start=1  
/product="capsid protein"  
/protein\_id="AAD27757.1"  
/db\_xref="GI:4689097"

CDS  
1. 4718  
/translation="MAADGYLPMWLEBNSEGREWMDLKPAPKPKANQOKODDGRG  
LVLPGYKYLGFENGLDKGEVPAVNAADAALEHDKAYDQQLKAGDNPYLRYNHADAEFQ  
RLQEDTSFGNLGRAVEQAKRVLPLGLVEGAKTAPGKRPVQSPQBPDSGIG  
KTQOPAKRLNFGQTDSESVDPQPLGEPATPAVGPPTMASGGGAPMADNNEGA  
DGVNAGSNHMDSTWLGDRVITSTRWALPTYNHLKYQISSASTGASNDNHFGY  
STPWGYFDENRFCHFSPPRQRLINNNWFRPKLNFLENQYKEVTINDQVTTIA  
NMLTSTVQVPSDSEYQLPYVLSAHOGLPPFPADVMIPOYQYJLNNGSQAVERSS  
FYCLEYFSPQMLRTGNFTFSYFEEVPHSSVYASOSLDRLNPLIDQLYLYLNRTQ  
NOSGAQNKDLIFSRGSPAGMSVQPKMLPQCYRQORVSKTKTDNNNSNFTWTGASK  
YNLNGRESIINPSTAMASHKDEDEKFPMSGVMIFGESAGASNTALDNMTDEEIE  
KATNPVATERFRTVAVNFQSSSTDPATGVDHAMGALPGMWQDRDYLYQGPIMAKIPH  
TDGHFHSPLMGFGFLKNPSPQLIKNTPVNPANPABFSATKFAFITQYSTGVQSVSE  
IEMELQKENSRRWNPVEQYTSNYSASANDFTVDNGLYTEPRPIGTRYLTRPL"  
BASE COUNT 1121 a 1393 c 1273 g 931 t  
ORIGIN

Query Match 100.0%; Score 4718; DB 14; Length 4718;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 4718; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCCCACTCCCTCTGCGCGCTCGCTCGTGGGGGCTGCGAACCAAGTCCGC 60  
DB 1 TTGCCCACTCCCTCTGCGCGCTCGCTCGTGGGGGCTGCGAACCAAGTCCGC 60  
QY 61 AGACGGCAGAGCTCTGCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120  
DB 61 AGACGGCAGAGCTCTGCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120  
QY 121 GCGCACTCCATCATTAGGGGTAATCGCGAAGCGCTCCCACTGCGCGCGCTGA 180  
DB 121 GCGCACTCCATCATTAGGGGTAATCGCGAAGCGCTCCCACTGCGCGCGCTGA 180  
QY 181 CGTAAATTACGTATAGGGGAGTGTCTGTATAGCTGTCACTGAGTGTCTTTGCGAC 240  
DB 181 CGTAAATTACGTATAGGGGAGTGTCTGTATAGCTGTCACTGAGTGTCTTTGCGAC 240  
QY 241 ATTTGCGACACCAAGTGCATTTAGGGTATATATGCGCGAGTGAAGCGAGATCTC 300  
DB 241 ATTTGCGACACCAAGTGCATTTAGGGTATATATGCGCGAGTGAAGCGAGATCTC 300

QY 301 CATTTGACCGCGAAATTGGAACGACGACGACCATGCCGCTTCTACGAGATCGTGAT 360  
DB 301 CATTTGACCGCGAAATTGGAACGACGACGACCATGCCGCTTCTACGAGATCGTGAT 360  
QY 361 CAAGTGCAGCGACCTGACGACGACCTCGCGGCAATTTGACTCGTTGTGAGCTG 420  
DB 361 CAAGTGCAGCGACCTGACGACGACCTCGCGGCAATTTGACTCGTTGTGAGCTG 420  
QY 421 GGTGCGCGAAGGAATGAGCTGCCCCCGGATTCTGACATGATCTGATCTGATGA 480  
DB 421 GGTGCGCGAAGGAATGAGCTGCCCCCGGATTCTGACATGATCTGATCTGATGA 480  
QY 481 GCAGGCAACCCCTGACCGGTGCGCGAAGCTGACGCGCACTTCTGCTCAATGCGCGC 540  
DB 481 GCAGGCAACCCCTGACCGGTGCGCGAAGCTGACGCGCACTTCTGCTCAATGCGCGC 540  
QY 541 CGTAGTAAGGCCCCCGGAGGCTTCTTCTTGTTCAGTTTCAGAGAAGGCGAGTCTACTT 600  
DB 541 CGTAGTAAGGCCCCCGGAGGCTTCTTCTTGTTCAGTTTCAGAGAAGGCGAGTCTACTT 600  
QY 601 CCACCTCCATATTCTGTGTGAGACCAAGGGGTCAATTCATGCTGTGGCGCTTCTCT 660  
DB 601 CCACCTCCATATTCTGTGTGAGACCAAGGGGTCAATTCATGCTGTGGCGCTTCTCT 660  
QY 661 GAGTCAGATTAGGACCAAGCTGTGTGACCACTTACCGCGGATCGAGCCGACCTTCC 720  
DB 661 GAGTCAGATTAGGACCAAGCTGTGTGACCACTTACCGCGGATCGAGCCGACCTTCC 720  
QY 721 CAACTGCTCGCGGTGACCAAGCGCTAATGCGCGCGAGGGGGAACAAGTGTGTGA 780  
DB 721 CAACTGCTCGCGGTGACCAAGCGCTAATGCGCGCGAGGGGGAACAAGTGTGTGA 780  
QY 781 CGAGTGTACATCCCACTACCTCTGCCCCAAGACTCAGCCCGAGCTGCAGTGGCGTG 840  
DB 781 CGAGTGTACATCCCACTACCTCTGCCCCAAGACTCAGCCCGAGCTGCAGTGGCGTG 840  
QY 841 GACTAATCATGAGAGGATATATAAGCGCTGTTGAACCTGCGCGGCAACCGCTCGT 900  
DB 841 GACTAATCATGAGAGGATATATAAGCGCTGTTGAACCTGCGCGGCAACCGCTCGT 900  
QY 901 GCGCGACGACCTGACCCACGTCAGCGACGACCGAGAGCAGAACAGGAATCTGAACCC 960  
DB 901 GCGCGACGACCTGACCCACGTCAGCGACGACCGAGAGCAGAACAGGAATCTGAACCC 960  
QY 961 CAATTCTGACCGCGCTGTATCCGCTCAAAAACTCCGCGCGCTATCATGAGCTGTGCG 1020  
DB 961 CAATTCTGACCGCGCTGTATCCGCTCAAAAACTCCGCGCGCTATCATGAGCTGTGCG 1020  
QY 1021 GTGCTGTGGAACCGGCGATCACTCCGAGAAGCAGTGCAGAGACGAGCGCTC 1080  
DB 1021 GTGCTGTGGAACCGGCGATCACTCCGAGAAGCAGTGCAGAGACGAGCGCTC 1080  
QY 1081 GTACATCTCTTCAACGCGCTTCCAACTCGCGGTCCAGATCAAGCGCGCTTGACAA 1140  
DB 1081 GTACATCTCTTCAACGCGCTTCCAACTCGCGGTCCAGATCAAGCGCGCTTGACAA 1140  
QY 1141 TGCGGCAAGATCATGCGCTGACCAAAATCCGCGCGCGACTTACCTGTAGGCCCGCTCC 1200  
DB 1141 TGCGGCAAGATCATGCGCTGACCAAAATCCGCGCGCGACTTACCTGTAGGCCCGCTCC 1200  
QY 1201 GCGCGCGGACATTAACAACCAACCGCATCTACCGCATCTTGAGCTGAACGGCTACGAAC 1260  
DB 1201 GCGCGCGGACATTAACAACCAACCGCATCTACCGCATCTTGAGCTGAACGGCTACGAAC 1260  
QY 1261 TGCTAGCGCGGCTCCGCTTCTCGGCTGGGCGCCAGAAAAGTTGCGGAAGCGCAACAC 1320  
DB 1261 TGCTAGCGCGGCTCCGCTTCTCGGCTGGGCGCCAGAAAAGTTGCGGAAGCGCAACAC 1320  
QY 1321 CATCTGCTGTTTGGGCGCGGCAACGCGGCAAGACCAACATCGCGGAAGCCATCGCCA 1380  
DB 1321 CATCTGCTGTTTGGGCGCGGCAACGCGGCAAGACCAACATCGCGGAAGCCATCGCCA 1380  
QY 1381 CGCGTGCCTTCTACGCGCTGCTCACTGGAACCAATGAGAACTTCCCTTCATGATTG 1440

Db 1381 CGCGTGCCTTCTACGGCTGCGTCACTGGAACCATGAGAACTTCCCTTCATGATTG 1440  
Qy 1441 CGTCGACAAGATGTGTATCTGTGGAGAGAGGCAAGATGACGGCCAAAGTCTGTGAGTC 1500  
Db 1441 CGTCGACAAGATGTGTATCTGTGGAGAGAGGCAAGATGACGGCCAAAGTCTGTGAGTC 1500  
Qy 1501 CGCCAAAGGCCATTCTCGGCGGAGCAAGGTGCGCGTGAGCCAAAGTGCAGAGTCTCCGC 1560  
Db 1501 CGCCAAAGGCCATTCTCGGCGGAGCAAGGTGCGCGTGAGCCAAAGTGCAGAGTCTCCGC 1560  
Qy 1561 CCAGATGACCCCAACCCCGTGATGTCACCTCCAACACCAACATGTGCGCGTGATTGA 1620  
Db 1561 CCAGATGACCCCAACCCCGTGATGTCACCTCCAACACCAACATGTGCGCGTGATTGA 1620  
Qy 1621 CGGAAACAGCAACCTTCGAGCACAGACCGGTTGAGAGCCGATGTTCAAAATTGA 1680  
Db 1621 CGGAAACAGCAACCTTCGAGCACAGACCGGTTGAGAGCCGATGTTCAAAATTGA 1680  
Qy 1681 ACTACCCCGCGTCTGAGCATGACTTTGGCAAGGTGACAAAGCAGAGAGTCAAGAGTT 1740  
Db 1681 ACTACCCCGCGTCTGAGCATGACTTTGGCAAGGTGACAAAGCAGAGAGTCAAGAGTT 1740  
Qy 1741 CTTCGCTGGGCGCAGGATCAGTCAACCGAGTGGCGCATGAGTTCTACGTAGAAAGG 1800  
Db 1741 CTTCGCTGGGCGCAGGATCAGTCAACCGAGTGGCGCATGAGTTCTACGTAGAAAGG 1800  
Qy 1801 TGGAGCCAAACAAAGACCCGCCCGATGACGCGGATAAAGCGAGCCCAAGCGGCGTG 1860  
Db 1801 TGGAGCCAAACAAAGACCCGCCCGATGACGCGGATAAAGCGAGCCCAAGCGGCGTG 1860  
Qy 1861 CCCCTCAGTGCAGATCCATGACGTGACGCGGAGAGAGCTCCGCTGGAATTGCGCA 1920  
Db 1861 CCCCTCAGTGCAGATCCATGACGTGACGCGGAGAGAGCTCCGCTGGAATTGCGCA 1920  
Qy 1921 CAGGTACCAAAACAATGTTCTGTCACGCGGCGATGTTCAAGATGCTTCCCTGCAA 1980  
Db 1921 CAGGTACCAAAACAATGTTCTGTCACGCGGCGATGTTCAAGATGCTTCCCTGCAA 1980  
Qy 1981 GACATGCGAGAGATGAATCAGAAATTTCAACATTTGCTTCAACGAGGAGAGAGACTG 2040  
Db 1981 GACATGCGAGAGATGAATCAGAAATTTCAACATTTGCTTCAACGAGGAGAGAGACTG 2040  
Qy 2041 TTCAAGATGCTTCCCGCGGTGTCAGAAATCTCAACCGGTGTCAGAAAGAGAGCTATCG 2100  
Db 2041 TTCAAGATGCTTCCCGCGGTGTCAGAAATCTCAACCGGTGTCAGAAAGAGAGCTATCG 2100  
Qy 2101 GAAACTCTGTCCATTCATCATCTGTGGGCGGGCTCCGAGATGCTTGTGCGGCTG 2160  
Db 2101 GAAACTCTGTCCATTCATCATCTGTGGGCGGGCTCCGAGATGCTTGTGCGGCTG 2160  
Qy 2161 CGATCTGTGCAACGTGACCTGATGACTGTGTTCTGAGCAATAATGACTTAACAG 2220  
Db 2161 CGATCTGTGCAACGTGACCTGATGACTGTGTTCTGAGCAATAATGACTTAACAG 2220  
Qy 2221 GTATGCTGCGATGTTATCTTCCAGATTGCTCGAGGCAACCTCTGAGGGCATTC 2280  
Db 2221 GTATGCTGCGATGTTATCTTCCAGATTGCTCGAGGCAACCTCTGAGGGCATTC 2280  
Qy 2281 GCGAGTGTGGGACTTGAAACCTGAGCCCGAAGCCCAAGCCCAAGCAAGCAAGAG 2340  
Db 2281 GCGAGTGTGGGACTTGAAACCTGAGCCCGAAGCCCAAGCCCAAGCAAGCAAGAG 2340  
Qy 2341 ACAGCGGCGGGGTCTGTGCTTCTGCTCAAGTACCTCGGACCTTCAACGGAATCG 2400  
Db 2341 ACAGCGGCGGGGTCTGTGCTTCTGCTCAAGTACCTCGGACCTTCAACGGAATCG 2400  
Qy 2401 ACAAGGGGAGCCCGTCAACGCGGAGAGAGCGGCGCTCGAGCAGCAAGGCTACG 2460  
Db 2401 ACAAGGGGAGCCCGTCAACGCGGAGAGAGCGGCGCTCGAGCAGCAAGGCTACG 2460  
Qy 2461 ACCAGAGCTCAAGCGGGTGACATCCGTAACCTGCGTATAACACGCGGAGCGGAGT 2520  
Db 2461 ACCAGAGCTCAAGCGGGTGACATCCGTAACCTGCGTATAACACGCGGAGCGGAGT 2520

Db 2461 ACCAGAGCTCAAGCGGGTGACATCCGTAACCTGCGTATAACACGCGGAGCGGAGT 2520  
Qy 2521 TTCAAGAGCGTCTGCAAGAGATACGTTCTTTGGGGCAACCTCGGGCAGAGTCTTCC 2580  
Db 2521 TTCAAGAGCGTCTGCAAGAGATACGTTCTTTGGGGCAACCTCGGGCAGAGTCTTCC 2580  
Qy 2581 AGGCCAAGAGCGGGTCTCGAACCCTCGGTCTGTGAGGAAGCGCTAAGACGCTC 2640  
Db 2581 AGGCCAAGAGCGGGTCTCGAACCCTCGGTCTGTGAGGAAGCGCTAAGACGCTC 2640  
Qy 2641 CTGGAAGAAACGTCCCGTAGACAGTCGCCACAAGAGCCAGACTCCTCCTCGGCATCG 2700  
Db 2641 CTGGAAGAAACGTCCCGTAGACAGTCGCCACAAGAGCCAGACTCCTCCTCGGCATCG 2700  
Qy 2701 GCAAGCAGGCCAGCAGCCCGCTAATAAAGAGACTCAATTTGGTCAAGCTGGCGACTCAG 2760  
Db 2701 GCAAGCAGGCCAGCAGCCCGCTAATAAAGAGACTCAATTTGGTCAAGCTGGCGACTCAG 2760  
Qy 2761 AGTCAGTCCCGATCCACAACCTCTCGGAGAACCTCCAGCAACCCCGCTGTGGAGC 2820  
Db 2761 AGTCAGTCCCGATCCACAACCTCTCGGAGAACCTCCAGCAACCCCGCTGTGGAGC 2820  
Qy 2821 CTACTCAATGAGCTTCAAGCGGTGGCGCACCAATGAGCAGATAACGAAGCGCGAGC 2880  
Db 2821 CTACTCAATGAGCTTCAAGCGGTGGCGCACCAATGAGCAGATAACGAAGCGCGAGC 2880  
Qy 2881 GAGTGGGTAATGCTCAGGAATTTGGCATTTGCGATTCCACATGCTGGGCGAGAGTCA 2940  
Db 2881 GAGTGGGTAATGCTCAGGAATTTGGCATTTGCGATTCCACATGCTGGGCGAGAGTCA 2940  
Qy 2941 TCACCAACAGCACCCGCACTGGGCGCTTCCACACTACATTAACCACTCTACAGCAAA 3000  
Db 2941 TCACCAACAGCACCCGCACTGGGCGCTTCCACACTACATTAACCACTCTACAGCAAA 3000  
Qy 3001 TCTCAGTGTCTTCAACGGGGGCGACCAAGCAACCACTTTCGCTACAGCACCCCT 3060  
Db 3001 TCTCAGTGTCTTCAACGGGGGCGACCAAGCAACCACTTTCGCTACAGCACCCCT 3060  
Qy 3061 GGGGTATTTTGAATTTCAACAGATTCACCTGCCACTTTTCAACAGTGGAGCGAC 3120  
Db 3061 GGGGTATTTTGAATTTCAACAGATTCACCTGCCACTTTTCAACAGTGGAGCGAC 3120  
Qy 3121 TCATCAACAACAATTTGGGATTCGGGCCCAAGAGACTCAACTTCAACATCC 3180  
Db 3121 TCATCAACAACAATTTGGGATTCGGGCCCAAGAGACTCAACTTCAACATCC 3180  
Qy 3181 AAGTCAAGAGGTACAGCAGATGATGGCGTCAACAACCATGCTAATACTTACCAGCA 3240  
Db 3181 AAGTCAAGAGGTACAGCAGATGATGGCGTCAACAACCATGCTAATACTTACCAGCA 3240  
Qy 3241 CGGTCAAGTCTTCTCGAGCTCGAGTACAGCTTCCGTAAGTCTCGGCTGCGCAC 3300  
Db 3241 CGGTCAAGTCTTCTCGAGCTCGAGTACAGCTTCCGTAAGTCTCGGCTGCGCAC 3300  
Qy 3301 AGGCTGCTCCCTCCGTTCCCGCGGAGCGTGTCAATGATTCGCAATACGGCTACCTGA 3360  
Db 3301 AGGCTGCTCCCTCCGTTCCCGCGGAGCGTGTCAATGATTCGCAATACGGCTACCTGA 3360  
Qy 3361 AGGCTGCTCCCTCCGTTCCCGCGGAGCGTGTCAATGATTCGCAATACGGCTACCTGA 3420  
Db 3361 AGGCTGCTCCCTCCGTTCCCGCGGAGCGTGTCAATGATTCGCAATACGGCTACCTGA 3420  
Qy 3421 CTTCTAGATGCTGAGAACGGGCAACAATTACCTTCAAGTACACCTTTGAGGAAGTGC 3480  
Db 3421 CTTCTAGATGCTGAGAACGGGCAACAATTACCTTCAAGTACACCTTTGAGGAAGTGC 3480  
Qy 3481 CTTTCAACAGCAGCTACGGCGACAGCCAGAGCTGGAACGGGCTGATGAATCTCTCATCG 3540  
Db 3481 CTTTCAACAGCAGCTACGGCGACAGCCAGAGCTGGAACGGGCTGATGAATCTCTCATCG 3540  
Qy 3541 ACCAATACCTGTATTACCTGAACAGAACTCAAAATCAGTCCGGAAGTGCCCAAAACAAG 3600  
Db 3541 ACCAATACCTGTATTACCTGAACAGAACTCAAAATCAGTCCGGAAGTGCCCAAAACAAG 3600



QY 3601 ACTTGCTGTTAGCCGTTGGGCTCTCCAGCTGGCATGTCTGTTACAGCCCAAAAACTGGCTAC 3660  
DB 3601 ACTTGCTGTTAGCCGTTGGGCTCTCCAGCTGGCATGTCTGTTACAGCCCAAAAACTGGCTAC 3660  
QY 3661 CTGGACCTGTTATCGGACGACGCGGCTTCTTAAAACAAAAAACAGACAACAACAGCA 3720  
DB 3661 CTGGACCTGTTATCGGACGACGCGGCTTCTTAAAACAAAAAACAGACAACAACAGCA 3720  
QY 3721 ATTTTACCTGACTGGTGTCTCAAAATATTAACCTCAATGGGCGTGAATCCATCAAC 3780  
DB 3721 ATTTTACCTGACTGGTGTCTCAAAATATTAACCTCAATGGGCGTGAATCCATCAAC 3780  
QY 3781 CTGGCACTGCTATGGCTCACAACAAGACGACGACCAAGTTCTTCCCATGAGCGGTG 3840  
DB 3781 CTGGCACTGCTATGGCTCACAACAAGACGACGACCAAGTTCTTCCCATGAGCGGTG 3840  
QY 3841 TCATGATTTTGGAAAAAGAGAGCGCGGAGCTTCAAACTGCATTGACATGTCATGA 3900  
DB 3841 TCATGATTTTGGAAAAAGAGAGCGCGGAGCTTCAAACTGCATTGACATGTCATGA 3900  
QY 3901 TTACAGACGAAGAGAAAAATTAAGCCACTAACCTGTGGCCACCGAAAGATTGGGACCG 3960  
DB 3901 TTACAGACGAAGAGAAAAATTAAGCCACTAACCTGTGGCCACCGAAAGATTGGGACCG 3960  
QY 3961 TGGCAGTCAATTTCCAGACGACGACGACGACGACGACGACGACGACGACGACGACG 4020  
DB 3961 TGGCAGTCAATTTCCAGACGACGACGACGACGACGACGACGACGACGACGACGACG 4020  
QY 4021 GAGCATTAACCTGGCATGCTGTGGCAAGATAGAGAGTGTACTGACAGGATCCCATTTGGG 4080  
DB 4021 GAGCATTAACCTGGCATGCTGTGGCAAGATAGAGAGTGTACTGACAGGATCCCATTTGGG 4080  
QY 4081 CCAAAATTCCTCACAAGATGAGACACTTTCACCCGCTCTCTTATGGGCGGCTTTGGAC 4140  
DB 4081 CCAAAATTCCTCACAAGATGAGACACTTTCACCCGCTCTCTTATGGGCGGCTTTGGAC 4140  
QY 4141 TCAAGAACCCGCTCTCAGATCTCATCAAAAAACGCGCTGTCTCCGCAATCTCCGG 4200  
DB 4141 TCAAGAACCCGCTCTCAGATCTCATCAAAAAACGCGCTGTCTCCGCAATCTCCGG 4200  
QY 4201 CGGAGTTTACGCTACAAAGTTTGTCTTCAATCAACCAATCTCCACAGACAAGTGA 4260  
DB 4201 CGGAGTTTACGCTACAAAGTTTGTCTTCAATCAACCAATCTCCACAGACAAGTGA 4260  
QY 4261 GTGTGAAATGAAATGGAGCTGCAGAAAGAAAAACGCAAGCGGTGAATCCGGAAGTGC 4320  
DB 4261 GTGTGAAATGAAATGGAGCTGCAGAAAGAAAAACGCAAGCGGTGAATCCGGAAGTGC 4320  
QY 4321 AGTACACATCCAAATTAAGCAAAATCTGCCAAGTTGATTTTACTGTGACAAATGAC 4380  
DB 4321 AGTACACATCCAAATTAAGCAAAATCTGCCAAGTTGATTTTACTGTGACAAATGAC 4380  
QY 4381 TTTATACTGAGCTCGCCCATTTGGACACCCGTTACCTTACCCCTGTAATTACGTG 4440  
DB 4381 TTTATACTGAGCTCGCCCATTTGGACACCCGTTACCTTACCCCTGTAATTACGTG 4440  
QY 4441 TTAATCAATAAACCGGTTGATTGCTTCAAGTTGAATTGGTCTCTCTCTTCTTATCT 4500  
DB 4441 TTAATCAATAAACCGGTTGATTGCTTCAAGTTGAATTGGTCTCTCTCTTCTTATCT 4500  
QY 4501 TATCGGTTACCATGTTATAGCTTACACATTAATGCTGTGTGCGCTTCGCGATTAAG 4560  
DB 4501 TATCGGTTACCATGTTATAGCTTACACATTAATGCTGTGTGCGCTTCGCGATTAAG 4560  
QY 4561 ACTTACGTCATCGGTTACCCCTAGTATGAGTGGCCACTCCCTCTCTGCGCGCTGCG 4620  
DB 4561 ACTTACGTCATCGGTTACCCCTAGTATGAGTGGCCACTCCCTCTCTGCGCGCTGCG 4620  
QY 4621 TCGCTCGTGGGCTGCGGACCAAAAGTCCGACAGCGGACGAGCTCTGCTGCGCGCC 4680  
DB 4621 TCGCTCGTGGGCTGCGGACCAAAAGTCCGACAGCGGACGAGCTCTGCTGCGCGCC 4680

QY 4681 CCACCGACGAGCGGCGCGACAGAGGAGTGGCAA 4718  
DB 4681 CCACCGACGAGCGGCGCGACAGAGGAGTGGCAA 4718  
RESULT 2 4683 bp DNA linear VRL 12-JAN-1998  
AF028704  
LOCUS  
DEFINITION Adeno-associated virus 6, complete genome.  
ACCESSION AF028704  
VERSION AF028704.1 GI:2766605  
KEYWORDS  
SOURCE adeno-associated virus 6.  
ORGANISM adeno-associated virus 6  
REFERENCE 1 (bases 1 to 4683)  
Rutledge, E.A., Halbert, C.L. and Russell, D.W.  
Infectious clones and vectors derived from adeno-associated virus  
(AAV) serotypes other than AAV type 2  
J. Virol. 72 (1), 309-319 (1998)  
JOURNAL  
MEDLINE 98080418  
PUBMED 9420229  
REFERENCE 2 (bases 1 to 4683)  
Rutledge, E.A. and Russell, D.W.  
Direct Submission  
Submitted (24-SEP-1997) Hematology, Univ. of Washington, Box  
357720, Seattle, WA 98195, USA  
FEATURES  
source  
1. 4683  
/organism="adeno-associated virus 6"  
/db\_xref="taxon:68558"  
/note="AAV6"  
320..2191  
/codon\_start=1  
/evidence=not experimental  
/product="nonstructural protein Rep78"  
/protein\_id="AAB95449.1"  
/db\_xref="GI:2766606"  
/translation="MPGFEIVIKVPSDLDEHLPGISDSFVNWVAEKWELEPPDSMD  
LNLIEQAPLVAEKLQDFLVQWRVSKAPEALFVQFEKESYFHLHIVETGVKS  
MVLGRFLSQIRDLVQTVIRGIEPTLPNMFATKTRNGAGGKVVDECIIPNYLLPK  
TQPELQWMTNMEETISACLNLAERKRLVAHDLTHVSQTOEQNKENLNPSDAVIRS  
KTSARMYELVGLVDRGITSEKOWIOEDQASYSFNAASNSRSQIKALDNAGKIML  
TKSAPDYLVPAPPADIKTNRIYRILELNGYDPAVAGSVFLGWAQKRFKNTIWLFG  
PATGKTINIAEAIHAAVPEYGCVMNTNENFPENDCVDKVIWMEEGKMTAKVYESAKA  
ILGSKVRVDQCKSSAQIDPTPIVITSNTNCAVIDGNTTFEHOQPLDRMFKFEL  
TRLEHDFGKVTQKEVEFEFRWAODHTEVAHEFYRKGGANKRAPADADKSEPKRA  
CPSVADPSTSDAEAPVDFADRYQNKCSRHAGMLQMLFPCKTCERNQNFNICFTHT  
RDCSECFPGVSESQPVVRKRTYRKLCAIHHLGRAPEIACSACDLVNVLDLDCVSEQ"  
2208..4418  
/codon\_start=1  
/evidence=not experimental  
/product="capsid protein VP1"  
/protein\_id="AAB95450.1"  
/db\_xref="GI:2766607"  
/translation="MAADGYLPDWLEDNLSEGREWMDLKGAPKPKANQKODDGRG  
LVLPGYKYLGPFGNGLDKGEPPVNAADAALEHDKAYDQQLKAGDNPYLRYNHADAEFOE  
RLQEDTSPFGNLRGAVFOAKKRVLEPFLVEGAKTAPGKKRPVEQSPQEPDSSGIG  
KTGOOPAKKRLNFGQDSESVDPDPLGEPATPAVGPPTMASGGGAPMADNNEGA  
DGVGNASGNWHCDSTWLGDRVITSTRTWALPTYNNHLYKQISSASTGSNDNHYFGY  
STPMGYDFNRFCHFSPRDMORLNNNWGFRPKRILFKLFNIQVKEVTNDGVTTIA  
NMLTSTVOVFSDEYQLPVYLSAHQGLPPPADVEMIPOYGYLTLANSSQAVGRSS  
FYCLEYFSPSOMLRITGNNTFSYTFEDVPFHSYSASQSLDRMLNPLIDQVLYLNRTQ  
NQSQAQNKDLFSRSGSPAGMSVQPKMWLPGPCYRQORVSKTKTDNNNSNFTWTGASK  
YNLNGRESTINPGTAMASHKDDKDFPMSGVIMFGKESAGASNTALDNVITDEEII  
KATNPVATERFGTVANLQSSSTDPAIGDVHNGALPGMWODRVDVYLQGIWAKIPIH  
TDGHPHSPLMGFGELKHPPIQILIKNTPVANPAPAEFATKPFASFTQYSTGQVSE  
IEMELXENSKRMNDEVOYTSNYAKSANVDFYVNNGLYTEPRPIGTRYLRLPL"

Query Match 90.3%; Score 4258; DB 14; Length 4683;  
Best Local Similarity 94.7%; Pred. No. 0;





QY 2158 CTGCGATCTGCTCAACGTGGACCTGGATGACTGTGTTCTGAGCAATAATGACTTAAC 2217  
 |||||  
 Db 2143 CTGCGATCTGCTCAACGTGGATCTGGATGACTGTGTTCTGAGCAATAATGACTTAAC 2202  
 |||||  
 QY 2218 CAGGTATGGCTGCCGATGGTTATCTTCAGATGGCTCGAGGACCAACCTCTCTGAGGGCA 2277  
 |||||  
 Db 2203 CAGGTATGGCTGCCGATGGTTATCTTCAGATGGCTCGAGGACCAACCTCTCTGAGGGCA 2262  
 |||||  
 QY 2278 TTCGAGTGTGGACTTGAACCTGGAGCCCGAAGCCCAAGCCCAACGCAAAAAGC 2337  
 |||||  
 Db 2263 TTCGAGTGTGGACTTGAACCTGGAGCCCGAAGCCCAAGCCCAACGCAAAAAGC 2322  
 |||||  
 QY 2338 AGGACGACGCGCGGGTCTGGTCTTCTGCTAGTACCTCGAACCTTCAACGGAC 2397  
 |||||  
 Db 2323 AGGACGACGCGCGGGTCTGGTCTTCTGCTAGTACCTCGAACCTTCAACGGAC 2382  
 |||||  
 QY 2398 TCGACAAGGGGAGCCCGTCAACGCGCGGACCGACGCGGCCCTCGAGCAGCAAGGCT 2457  
 |||||  
 Db 2383 TCGACAAGGGGAGCCCGTCAACGCGCGGATGCAAGCGGCCCTCGAGCAGCAAGGCT 2442  
 |||||  
 QY 2458 AGCACCAGCAGCTCAAAAGGGGTGACAACTCCGTACTGCGGTATTAACACGCGCAGCGCG 2517  
 |||||  
 Db 2443 AGCACCAGCAGCTCAAAAGGGGTGACAACTCCGTACTGCGGTATTAACACGCGCAGCGCG 2502  
 |||||  
 QY 2518 AGTTTCAAGAGCGTCTGCAAGAAATACGTCTTTGGGGGCAACCTCGGGCAGCAGTCT 2577  
 |||||  
 Db 2503 AGTTTCAAGAGCGTCTGCAAGAAATACGTCTTTGGGGGCAACCTCGGGCAGCAGTCT 2562  
 |||||  
 QY 2578 TCCAGGCCAAGAGCGGTTCTCGAACCTCTCGGTCTGTTGAGGAAGCGCTAAGACGG 2637  
 |||||  
 Db 2563 TCCAGGCCAAGAGAGGTTCTCGAACCTTTTGGTCTGTTGAGGAAGGTCTAAGACGG 2622  
 |||||  
 QY 2638 CTCCTGGAAGAAAGTCGCGTAGAGCAGTCCGCCAAGAGCCAGACTCCTCTCGGGCA 2697  
 |||||  
 Db 2623 CTCCTGGAAGAAAGTCGCGTAGAGCAGTCCGCCAAGAGCCAGACTCCTCTCGGGCA 2682  
 |||||  
 QY 2698 TCGGCAAGACAGGCCAGCAGCCCGCTAAAAAGACTCAATTTTGTCAAGTGGCGACT 2757  
 |||||  
 Db 2683 TCGGCAAGACAGGCCAGCAGCCCGCTAAAAAGACTCAATTTTGTCAAGTGGCGACT 2742  
 |||||  
 QY 2758 CAGAGTCAGTCCCGATCCACAACCTCTCGAGAACCTCCAGCAACCCCGCTGCTGTG 2817  
 |||||  
 Db 2743 CAGAGTCAGTCCCGATCCACAACCTCTCGAGAACCTCCAGCAACCCCGCTGCTGTG 2802  
 |||||  
 QY 2818 GACCTACTACAATGGCTTCAGGCGGTGGCGCACCAATGGCAGACAATTAAGAGGCGCG 2877  
 |||||  
 Db 2803 GACCTACTACAATGGCTTCAGGCGGTGGCGCACCAATGGCAGACAATTAAGAGGCGCG 2862  
 |||||  
 QY 2878 ACGAGTGGTAATGCTCAGGAATTTGCATTGGATTCCACATGGCTGGCGCAGAG 2937  
 |||||  
 Db 2863 ACGAGTGGTAATGCTCAGGAATTTGCATTGGATTCCACATGGCTGGCGCAGAG 2922  
 |||||  
 QY 2938 TCATCACCACGACACCCGCACTGGGCTTGCCCACTACATTAACCACTCTACAAGC 2997  
 |||||  
 Db 2923 TCATCACCACGACACCCGCACTGGGCTTGCCCACTACATTAACCACTCTACAAGC 2982  
 |||||  
 QY 2998 AAATCTCAGTCTTCAAAGGGGGCCAGCAACCACTACTTGGCTACAGCAACC 3057  
 |||||  
 Db 2983 AAATCTCAGTCTTCAAAGGGGGCCAGCAACCACTACTTGGCTACAGCAACC 3042  
 |||||  
 QY 3058 CCTGGGGTATTTGATTTCAACAGATTCCACTGCACTTTTCAACCACTGGCAGC 3117  
 |||||  
 Db 3043 CCTGGGGTATTTGATTTCAACAGATTCCACTGCACTTTTCAACCACTGGCAGC 3102  
 |||||  
 QY 3118 GACTCATCAACAACAATTGGGGATTCCGGCCCAAGAGACTCAACTCAACTCTTCAACA 3177  
 |||||  
 Db 3103 GACTCATCAACAACAATTGGGGATTCCGGCCCAAGAGACTCAACTCAACTCTTCAACA 3162  
 |||||  
 QY 3178 TCCAAGTCAAGAGGTCAAGCAGATGATGGCGTCAACCACTCGTAATACTTACCA 3237  
 |||||  
 Db 3163 TCCAAGTCAAGAGGTCAAGCAGATGATGGCGTCAACCACTCGTAATACTTACCA 3222  
 |||||  
 QY 3238 GCAAGGTTCAAGTCTTCTCGGACTCGGAGTACCACTTCCGTACGCTCGGCTCGCGC 3297  
 |||||

Db 3223 GCAAGGTTCAAGTCTTCTCGGACTGGAGTACCAAGTTGCCGTACGTCTCGGCTCGCGC 3282  
 |||||  
 QY 3298 ACCAGGGTGCCTCCCTCCGTTCCGGCGGACGTGTTCATGATTCGCAATACGGCTACC 3357  
 |||||  
 Db 3283 ACCAGGGTGCCTCCCTCCGTTCCGGCGGACGTGTTCATGATTCGCAATACGGCTACC 3342  
 |||||  
 QY 3358 TGACGCTCAACAATGGCAGCCAGCCGTGGAGCTTCATCTTTTACTGCTGGAATATT 3417  
 |||||  
 Db 3343 TACGCTCAACAATGGCAGCCAGCAGTGGAGCGGTATCTTTTACTGCTGGAATATT 3402  
 |||||  
 QY 3418 TCCCTTCTCAGATGCTGAGAACGGGCAACAATTACCTTCACTACCTTGAGGAG 3477  
 |||||  
 Db 3403 TCCCATCGCAGATGCTGAGAACGGGCAATACTTACCTTCACTACCTTGAGGAG 3462  
 |||||  
 QY 3478 TGCCCTTCCACAGCAGCTACGCGCAGCAGCAGCCGTGAGCCGCTGATGAATCTCTCA 3537  
 |||||  
 Db 3463 TGCCCTTCCACAGCAGCTACGCGCAGCAGCAGCCGTGAGCCGCTGATGAATCTCTCA 3522  
 |||||  
 QY 3538 TCGACCAATACCTGTATTACCTGAACAGAACTCAAAATCAGTCCGGAAGTGCCCAAAACA 3597  
 |||||  
 Db 3523 TCGACCAATACCTGTATTACCTGAACAGAACTCAGAACTCAGTCCGGAAGTGCCCAAAACA 3582  
 |||||  
 QY 3598 AGGACTTGTGTTTAGCCGTGGTCTCCAGCTGCGATGTCTGTCAAGCCCAAAACTGCG 3657  
 |||||  
 Db 3583 AGGACTTGTGTTTAGCCGTGGTCTCCAGCTGCGATGTCTGTCAAGCCCAAAACTGCG 3642  
 |||||  
 QY 3658 TACCTGACCCCTGTATTACGCGCAGCAGCGGCTTTCTAAACAACAAACAGACACACACACA 3717  
 |||||  
 Db 3643 TACCTGACCCCTGTATTACGCGCAGCAGCGGCTTTCTAAACAACAAACAGACACACACACA 3702  
 |||||  
 QY 3718 GCAATTTTACCTGAGCTGTGTCTTCAAAATATACTCAATGGGCGTGAATCCATCATCA 3777  
 |||||  
 Db 3703 GCAATTTTACCTGAGCTGTGTCTTCAAAATATACTCAATGGGCGTGAATCCATCATCA 3762  
 |||||  
 QY 3778 ACCCTGCACTGCTATGGCTCACAACAAGACGAGCAAGCAAGTCTTCCCATGAGCG 3837  
 |||||  
 Db 3763 ACCCTGCACTGCTATGGCTCACAACAAGACGAGCAAGCAAGTCTTCCCATGAGCG 3822  
 |||||  
 QY 3838 GTGTCAATTTTGGAAAAAGAGCGCCGAGCTTCAACACTGATTGGACAATGTCA 3897  
 |||||  
 Db 3823 GTGTCAATTTTGGAAAAAGAGCGCGGAGCTTCAACACTGATTGGACAATGTCA 3882  
 |||||  
 QY 3898 TGATTACAGACGAAGAGGAATTAAGCCACTAACCTGTGGCCACCGAAGATTGGGA 3957  
 |||||  
 Db 3883 TGATTACAGACGAAGAGGAATTAAGCCACTAACCCGTGGCCACCGAAGATTGGGA 3942  
 |||||  
 QY 3958 CCGTGGCAGTCAATTTCCAGAGCAGCAGCAAGACCCCTGCGACCGGAGATGTGCATGCTA 4017  
 |||||  
 Db 3943 CCGTGGCAGTCAATTTCCAGAGCAGCAGCAAGACCCCTGCGACCGGAGATGTGCATGCTA 4002  
 |||||  
 QY 4018 TGGGAGCATTACTGCGCATGGTGGCAAGATAGAGACGTGTACTGACAGGGTCCCATTT 4077  
 |||||  
 Db 4003 TGGGAGCATTACTGCGCATGGTGGCAAGATAGAGACGTGTACTGACAGGGTCCCATTT 4062  
 |||||  
 QY 4078 GGGCCAAAATTCCTCACACAGATGGAACATTTACCCGCTCTCTTATGGGCGGCTTTG 4137  
 |||||  
 Db 4063 GGGCCAAAATTCCTCACACAGATGGAACATTTACCCGCTCTCTTATGGGCGGCTTTG 4122  
 |||||  
 QY 4138 GACTCAAGAACCCGCTCTCAGATCCTCATCAAAAAACAGCGCTGTCTCTGGAATCTC 4197  
 |||||  
 Db 4123 GACTCAAGAACCCGCTCTCAGATCCTCATCAAAAAACAGCGCTGTCTCTGGAATCTC 4182  
 |||||  
 QY 4198 CGGCGGATTTTCAAGTCAAGATTTGCTTCACTTCACTCAACCAATACTCCACAGGACAAG 4257  
 |||||  
 Db 4183 CGGCGGATTTTCAAGTCAAGATTTGCTTCACTTCACTCAACCAATACTCCACAGGACAAG 4242  
 |||||  
 QY 4258 TGAGTGGAAAAATTGAATGGAGCTGCAGAAAGAAAAACAGACGCGCTGGAATCCGGAAG 4317  
 |||||  
 Db 4243 TGAGTGGAAAAATTGAATGGAGCTGCAGAAAGAAAAACAGACGCGCTGGAATCCGGAAG 4302  
 |||||  
 QY 4318 TGAGTACACATCCAAATTATGCAAAATCTGCCAACGTTGATTTTACTGTGACAAACAATG 4377  
 |||||



Db 4303 TGCAATATACATCTAATAATGCAAAATCTGCCAACGTTGATTTCACCTGTGACAAACAATG 4362  
QY 4378 GACTTATATCTGAGCCTCGCCCCCATTTGGCAACCCGTTACCTTACCCGTCCTGTAATATAC 4437  
Db 4363 GACTTATATCTGAGCCTCGCCCCCATTTGGCAACCCGTTACCTTACCCGTCCTGTAATATG 4422  
QY 4438 GTGTTATCAATAAACCAGTTGATTGTTTCAAGTTGAACCTTTGCTCTCTGCTTCTTA 4497  
Db 4423 GTGTTATCAATAAACCAGTTGATTGTTTCAAGTTGAACCTTTGCTCTCTGCTTCTTA 4482  
QY 4498 TCTTATC-GGTTACCATGTTATAGCTTACACATTACTGCTTGTGCGCTTCCGATA 4556  
Db 4483 TCTTATCTGCTACCATAGCAACCCGTTACACATTACTGCTTGTGCGCTTCCGATA 4540  
QY 4557 AAAGACTTACGTCATCGGGTTACCCCTAGTATGATGAGTTGCCCACTCCCTCTGCGCGC 4616  
Db 4541 -----ATACCTTAGTATGAGTTGCCCACTCCCTCTATGCGCGC 4581  
QY 4617 TCGCTCGCTCGGTGGGCGCTGCGGACCAAGTCCGACAGCGGACAGCTCTGCTCTGCC 4676  
Db 4582 TCGCTCGCTCGGTGGGCGCGGACAGAGAGCTCTGCGCTGCGGACCTTGTGCTCGCA 4641  
QY 4677 GGCCCAACCGAGCGAGCGCGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4718  
Db 4642 GGCCCAACCGAGCGAGCGCGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4683

RESULT 3

AVU48704 4726 bp DNA linear VRL 15-JUL-1996  
LOCUS Adeno-associated virus 3 nonstructural protein and capsid protein  
DEFINITION genes, complete cds, and complete genome.

ACCESSION U48704  
VERSION U48704.1 GI:1408467  
KEYWORDS Adeno-associated virus 3 strain=3H.  
SOURCE adeno-associated virus 3  
ORGANISM Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.

REFERENCE 1 (bases 1 to 4726)  
AUTHORS Muramatsu, S., Mizukami, H., Young, N.S. and Brown, K.E.  
TITLE Nucleotide sequencing and generation of an infectious clone of  
adeno-associated virus 3

JOURNAL Virology 221 (1), 208-217 (1996)  
MEDLINE 96266430  
PUBMED 8661429

REFERENCE 2 (bases 1 to 4726)  
AUTHORS Muramatsu, S. and Brown, K.E.  
TITLE Direct Submission

JOURNAL Submitted (06-FEB-1996) Shin-ichi Muramatsu, Hematology Branch,  
NHBLI/NIH, 9000 Rockville, MD 20892, USA  
FEATURES location/Qualifiers

source 1. 4726  
/organism="adeno-associated virus 3"  
/strain="3H"  
/db\_xref="taxon:46350"  
318..2192  
/codon\_start=1  
/product="nonstructural protein"  
/protein\_id="AAC55048.1"  
/db\_xref="GI:1408467"

translation="MPGEYELVLPSPDLDERLPEISNSFVNWVAEKENDVPDSDMD  
PNLEQAPLTVAEKLOREPLVEMRVSKEPEALFEVQFEKGETYFHLVLETTIGVKS  
MVGGRVSOIKEKLVIRIYRIVGEPOLPNMFAVTKTRNGAGGKNVDDCTIPNLLPK  
TPELOAWMTNMDOYLASACINLAERKLVASHLTHVSQTOEONKONPNPSDAVIRS  
KTSARYMELVGMVDRGITSEKOWIQEDQASTISFNAASNSRSQIKALDNASKIMSL  
TKTAPDYLVGNSNPEDITKRIYQILELNGYDPOVAASVFLGAAOKKFGKNTIWLFG  
PATTKNTINAEIAHAVPFYGCVMNTNENFPENDCVDMVIMWEEGKMTAKVESAKA  
ILGSKVRVDOKCKSSAOIEPTPIVTSNTNMCVIDGNSSTFEHQPLDRMFEFEL  
TRRLDHPGKVTQKQEVYDFRWDSDHVTDAHEFYVRKGAKKRPASNDADVSEPKRE  
CTSLAQPTSDAEAPADYADRYONKCSRHVGMNMLFPCKTCERMNQISNVCFTHGQR  
DCGECFPGMESQPVSVVKKKTYQKLCPIHHLIGRAPELIACSACDLANVLDDCVSEQ  
2209..4419

/codon\_start=1  
/product="capsid protein"  
/protein\_id="AAC55049.1"  
/db\_xref="GI:1408469"  
/translation="MAADGYLPDWLEDNLSREIREWMAKPGVPQPKANQHQDNRRG  
LVLPGYKYLPGENGIDKGEPEVNEADAAALHDKAYDQOLKAGDNPYLKYNHADAFOE  
RLQEDTSFGNLRGAVFQAKRILEPLGVBEAAKTAIPKKGAVDQSPQPDSSSGVG  
KSGQPKARKRLNFGQTDSESVDPQPLGEPAPAPTSLSGNTAGSDVGAADNHYEYS  
DVGNSSGNMHCDOSMLGDRVITSTRTALPTYNHLYKQISSQSGASNDNHYEYS  
TPMGYFDENRPFHCSPRDWRLLNNWGFPRKLSFKLFNIQVRGVTQNDGTTTIAN  
NLSTVQVFTDSRYQLPYVLGSAHQGLPPPADVFMVPOGYLTLNNSQAVGRSSF  
XCLEYFSPQMLRTGNNEOFSTFEDVPHSSYASOSLDRLMNP.LIDQYLYLNRTOG  
TTSCTNQSRLIFSQAQPSQMSLQARNWLPQPCYRQRLSKTANDNNNSNFPWTAASK  
YHLNGRDSLVPNPBPAMASHKDDDEKFPFMHGNLI.FGKEGTASNAELDNVMTDEEBI  
RTNPNVATEQYGVANLQSSNTAPPTVNHQALPGMWQDRDVLQGP1WAKIPH  
TDGHEFSPPLMGFRGLKHPPOIMIKNTPVANPPTTSPAKFASFIQYSTQVSE  
IEWELQKENSXKWNPEIQYTSNYSNVNDFVDTNNGVSEPRPIGTRYLIRNL"

BASE COUNT 1200 a 1295 c 1251 g 980 t  
ORIGIN  
Query Match 68.1%; Score 3211.6; DB 14; Length 4726;  
Best Local Similarity 81.4%; Pred. No. 0;  
Matches 3865; Conservative 0; Mismatches 829; Indels 56; Gaps 10;

QY 1 TTGCCCACTCCCTCTCTGCGCGCTCGCTCGCTCGGTGGGCGCTCGGACCAAGATCCGC 60  
Db 1 TTGCCCACTCCCTCTATGCGCACTCGCTCGCTCGGTGGGCGCTCGGACCAAGATCCGC 60  
QY 61 AGACGGCAGAGCTCTGCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGTG 120  
Db 61 AGACGGCAGCTGCTTTGACGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGTG 120  
QY 121 GGCAACTCCATCACTAGGGGTAATCGGAAGCGCTCCACGCTGCCGCGTACGCGCTGA 180  
Db 121 GGCAACTCCATCACTAGAGTATGCGAGTACGTACGTACGTACGTACGTACGTACGTAC 180  
QY 181 CGTAAATTACGTCATAGGGGAGTGTCTGTATATGCTGTACGTGAGTGTGTCGAC 240  
Db 181 CGCCTACCA-----GCTGCGTACGAGTCAAGTGA-CCCTTTGCGAC 222  
QY 241 ATTTGCGACACCACTGCGCATTTAGGGTATATATGCGCGAGTGAGCG-AGCAGATCT 299  
Db 223 AGTTGCGACACCACTGCGCGCGCTGAGGGTATATATCTCGAGTGAGCGAGCAGAGCT 282  
QY 300 CCAATTTGACCGCGAAATTTGAACGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGTGA 359  
Db 283 CCAATTTGACCGCGAAATTTGAACGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGTGTCC 342  
QY 360 TCAAGTGCGCGAGCGAGCTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGTCT 419  
Db 343 TGAAGTGCGCGAGCTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGTCTTA 402  
QY 420 GCGTGCGCGAGAGAGATGGAGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 479  
Db 403 GCGTGCGCGAGAGAGATGGAGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 462  
QY 480 AGCAGGCAACCCCTGACCGTGGCGCGAGAGCTGACGCGCGCACTTCTGCTCAATGCGCGC 539  
Db 463 AGCAGGCAACCCCTGACCGTGGCGCGAGAGAGCTTCAAGCGCGAGTCTCTGAGAGTGGCGC 522  
QY 540 GCGTGAAGTAAAGCT 599  
Db 523 GCGTGAAGTAAAGCT 582  
QY 600 TCCACCTCCATATTTCTGTTGAGAGCAGCAGCGGGTCAAAATCCATGTGTGCGCGCGCTCC 659  
Db 583 TCCACCTGACGCTGCTATGAGACCATCGGGGTCAAAATCCATGTGTGCGCGCGCTACG 642  
QY 660 TGAATCAGATTAGGAGCAAGCTGTGAGACCATCTACCGCGGATGAGCGGACCGCTGC 719  
Db 643 TGAGCCAGATTAAAGAGAGAGCTGTGAGCCGCGCATCTACCGCGGATGAGCGGACCGCTTC 702  
QY 720 CCAACTGTTGCGCGTGAACCAAGACGCGTAAATGCGCGCGCGGAGGAGGAGCAAGGTGTGG 779

Db 703 CGAACTGGTTCGGCGGTGACCAAAAACCGGAAATGGCGCCGGGGCGGGGAACAAGGTGGTGG 762  
QY 780 ACAGTGTCTACATCCCACTACTACCTCTGCGCAAGACTCAGCCCGAGCTGCAGTGGCGGT 839  
Db 763 ACGACTGTCTACATCCCACTACTACCTCTGCGCAAGACTCAGCCCGAGCTGCAGTGGCGGT 822  
QY 840 GGACTTAACATGAGAGATATATTAAGCGCTGTGTTGAACCTGGCCGAGCGCAACCGGCTCG 899  
Db 823 GGACTTAACATGAGAGATATTAAGCGCTGTGTTGAATCTCGCGGAGCGTTAAACGGCTCG 882  
QY 900 TGGCGCAGCAGCTGACCCACGTCAGCCAGACCCAGAGCAGAACAGAGAAATCTGAACC 959  
Db 883 TGGCGCAGCAGCTGACCCACGTCAGCGAGAGCAGAGCAGAACAGAGAAATCAGAACC 942  
QY 960 CCAATTCTGACGCGCTGTCTATCCGCTCAAAAACCTCCGCGCTACATGAGCTGTGCG 1019  
Db 943 CCAATTCTGACGCGCGCTGTCTATCAGGTCAAAAACCTCAGCCAGGTACATGAGCTGTGCG 1002  
QY 1020 GGTGGCTGTGGACCGGGGCATCACCCTCCGAGAAAGCAGTGGATCCAGAGAACCAAGCCT 1079  
Db 1003 GGTGGCTGTGGACCGGGGCATCAGTCAAGAAAAGCAATGGATTGAGAGAACCAAGCCT 1062  
QY 1080 CGTACATCTCTTCAACGCGCTTCCAACTCGCGGTCCAGATCAAGCGCGCTCTGACA 1139  
Db 1063 CGTACATCTCTTCAACGCGCTTCCAACTCGCGGTCCAGATCAAGCGCGCTCTGACA 1122  
QY 1140 ATGCCGCAAGATCATGCGCTGACCAAAATCGCGCCGACTACCTGTAGAGCCCGCTC 1199  
Db 1123 ATGCCCTCAAGATCATGAGCCTGACAAAGACGCGCTCCGAGCTACCTGTGGGACGAAACC 1182  
QY 1200 CGCCCGCGGACATTAACAACCAACCGCATCTACCTCTGAGCTGAACGCTACGAAC 1259  
Db 1183 CGCCGAGGACATTAACAATAATCGATCTACCAATCCTGAGCTGAACGCGGTACGATC 1242  
QY 1260 CTGCCTACGCGCGCTCCGTCTTCTCGGCTGGGCCAGAAAAGTTGCGGAAGCGCAACA 1319  
Db 1243 CGCAGTACGCGCGCTCCGTCTTCTCGGCTGGGCCAGAAAAGTTGCGGAAGAGAACACA 1302  
QY 1320 CCATCTGGCTGTTGGGCCCGGCGCACCAAGCAATCGCGGAAGCCATCGCCC 1379  
Db 1303 CCATCTGGCTCTTTGGGCCCGGCGCACCAAGCAATCGCGGAAGCCATCGCCC 1362  
QY 1380 ACGCCGTGCTTCTTCTAAGCGCTGCTCAACTGGAACCAATGAGACTTCTCTCAATGATT 1439  
Db 1363 ACGCCGTGCTTCTTCTAAGCGCTGCTCAACTGGAACCAATGAGACTTCTCTCAATGATT 1422  
QY 1440 GCGTGAACAAGATGATGATCTGTGGAGAGAGGCAAGATGACGCGCAAGTCTGTGAGT 1499  
Db 1423 GCGTGAACAAGATGATGATCTGTGGAGAGAGGCAAGATGACGCGCAAGTCTGTGAGT 1482  
QY 1500 CGGCAAGGCAATCTCGCGCGGACGAAGGTGCGGTGGAACCAAAAGTCAAGTCTGCCG 1559  
Db 1483 GCGCAAGGCAATCTCGCGCGGAGCAAGGTGCGGTGGAACCAAAAGTCAAGTCTGCCG 1542  
QY 1560 CCGAGATCGACCCCAACCCCGGTGATCGTCACTCTCAACACCAACATGTGCGCGGTGATTG 1619  
Db 1543 CCGAGATCGACCCCAACCCCGGTGATCGTCACTCTCAACACCAACATGTGCGCGGTGATTG 1602  
QY 1620 ACGGGAACAGACCACTTTCAGACACGAGCGCGTTGACAGACCGGATGTTCAATTTG 1679  
Db 1603 ACGGGAACAGACCACTTTCAGACATCAGACGCGCTGCAAGACCGGATGTTTGAATTTG 1662  
QY 1680 AACTCACCCGCGCTGTGAGCATGACTTTGGCAAGGTGACAAAGCAGAGATCAAAAGAT 1739  
Db 1663 AACTCACCCGCGCTGTGAGCATGACTTTGGCAAGGTGACAAAGCAGAGATCAAAAGAT 1722  
QY 1740 TCTTCCGTGGCGCAGATCAAGTCAAGCGAGGTGGCGCATGATTCTACGTCAAGAAAG 1799  
Db 1723 TTTTCCGTGGCGCTTCCGATCAAGTCAAGTGGCTCATGAGTTCTACGTCAAGAAAG 1782  
QY 1800 GTGAGCCAACAAGAACCCCGCCGATGACGGGATTAAGCGAGCCCAAGCGGCGCT 1859

Db 1783 GTGAGCTAAGAAACGCCCCCGCTCCAATGACCGCGATGTAAGCGAGCCAAACCGGAGT 1842  
QY 1860 GCCCTCAGTCCGGGATCCATCGACGTGACAGCGGAAAGAGCTCCGGTGACTTTGCCG 1919  
Db 1843 GCAGTCACTTCCGAGCCGACAAAGTCAAGACGCGGA---AGCACCGCGGACTACGCGG 1899  
QY 1920 ACAGGTACCAAAACAATGTTCTGTCACGCGGCGCATGCTTCAGATGCTGTTCCCTGCA 1979  
Db 1900 ACAGGTACCAAAACAATGTTCTGTCACGCGGCGCATGATGCTTTTCCCTGTA 1959  
QY 1980 AGACATCGGAGAGATGAATCAGAATTTCAACATTTGCTTCACGCAAGGAGCAGAGACT 2039  
Db 1960 AAACATGCGAGAGATGAATCAATTTCCAATGTCTGTTTACGCATGTGTCAAAGAGACT 2019  
QY 2040 GTTCAGATGCTTCCCGCGCGTTCAGAAATCTAAC-----CGGTGTCAGAAAGAGGA 2093  
Db 2020 GTGGGGAATGCTTCCCTGGAATGTCAGAAATCTAACCCGTTTCTGTCTCAAAAAGAGGA 2079  
QY 2094 CGTATCGAAAATCTGTGCAATTCATCTGCTGGCGCGGCTCCGAGATTGCTTGT 2153  
Db 2080 CTATCAGAAAATCTGTGCAATTCATCTGCTGGAGGAGCAACCGAGATTGCTGTT 2139  
QY 2154 CGGCTGCGATCTGTCAACGTGGAACCTGATGACTGTGTTCTGAGCAATTAATGACTT 2213  
Db 2140 CGGCTGCGATTTGGCCAAATGTGACTTGGATGACTGTGTTCTGAGCAATTAATGACTT 2199  
QY 2214 AAACGAGTATGCTGCCGATGTTATCTTCCAGATTGGCTCGAGGACAACTCTGTAG 2273  
Db 2200 AAACGAGTATGCTGCTGCTGACGTTATCTTCCAGATTGGCTCGAGGACAACTTCTGAA 2259  
QY 2274 GGCATTGCGGAGTGTGGGACTTGAACCTTGAGGCCCGGAGCCCAAGCCCAAGCAGCA 2333  
Db 2260 GGCATTGCTGAGTGTGGGCTTGAACCTTGAGTCCCTCAACCCCAAGCAGCAACCA 2319  
QY 2334 AAGCAGACGACGCGCGGCTGTGCTTCTTCTGCTTCAAGTAACTCTCGAACCTTCAAC 2393  
Db 2320 CACGAGCAACCGTGGGGCTTGTGCTTCCGGTTACAAATACCTCGAACCCGTTAC 2379  
QY 2394 GGAATCGACAAGGGGAGCCCGTCAACGCGGAGCAGCGGCGCTGAGCAGCAGCAAG 2453  
Db 2380 GGAATCGACAAGGAGCCCGTCAACGAGCGGAGCAGCGGCGCTGAGCAGCAGCAAA 2439  
QY 2454 GCTACGACGACGACTCAAAAGCGGTGACAAATCCGTACTGCGGTATTAACACGCGGAC 2513  
Db 2440 GCTACGACGACGACTCAAAAGCGGTGACAAACCGTACTCAAGTACAAACGCGGAC 2499  
QY 2514 GCGGATTTACGAGAGCGTCTGCAAGAAAGATAGTCTTTTGGGGCAACCTCGGGCAGCA 2573  
Db 2500 GCGGATTTACGAGAGCGTCTTCAAGAAAGATAGTCTTTTGGGGCAACCTTGCGCAGCA 2559  
QY 2574 GTCTTCAAGGCAAGAGCGGTTCTGAACTCTCGGTCTGTTGAGGAAGCGGCTAAG 2633  
Db 2560 GTCTTCAAGGCAAGAGGATCTTGAAGCTCTTGGTCTGTTGAGGAAGCAGCTTAA 2619  
QY 2634 ACGGCTCTGAAAAGAAAGTCCGGTGAAGCAGTCCGCAAGAGCCAGACTCTCTCTCG 2693  
Db 2620 ACGGCTCTGAAAAGAGGCGGTGATGATCAGTCTCTCAGGAAACCGGACTCATCATCT 2679  
QY 2694 GGCATCGGCAAGACAGGCGCAGACCCCGTAAAAAGAGACTCAATTTTGTGACAGTGGC 2753  
Db 2680 GGTGTGGCAAAATCGGCAAGAGCGCTGCAAGAAAAAGCTAAATTTGCGTCAAGTGA 2739  
QY 2754 GACTCAGATCAGTCCCGGATCCACAACCTCTCGAGAACTTCCAGCAACCCCGCTGCT 2813  
Db 2740 GACTCAGATCAGTCCCGGATCCACAACCTCTCGAGAACTTCCAGCAACCCCGCAGAGT 2799  
QY 2814 GTGGGACTACTACATGCTTCAAGCGCGGTGGCGCACCAATGGCAGACATAAGAGAGG 2873  
Db 2800 TTGGGACTTAATACATGCTTCAAGCGCGGTGGCGCACCAATGGCAGACATAAGAGAGT 2859  
QY 2874 GCGGAGGAGTGGGTAATGCTTCAAGAAATTTGGCATTTGCAATTCACATGCTGGCGAC 2933  
Db 2860 GCGGAGGAGTGGGTAATTTCTCAGGAAATTTGCAATTTGCAATTCACATGCTGGCGAC 2919

QY 2934 AGAGTCATCAACCAACGACCCGCACTGGGCTTGCCCACTACATAACCACCTCTAC 2993  
Db 2920 AGAGTCATCAACCAACGACCAAGAACTGGGCTTGCCCACTACATAACCACCTCTAC 2979  
QY 2994 AAGCAATCTCCAGTGTCTCAACGGGGCCAGCAACGACCACTACTTGGCTACAGC 3053  
Db 2980 AAGCAATCTCCAGCAATCA--GGAGCTTCAACGACCACTACTTGGCTACAGC 3036  
QY 3054 ACCCTGGGGGATTTTGAATTCACAGATTCACCTGCACTTTTCAACCACTGACTGG 3113  
Db 3037 ACCCTGGGGGATTTTGAATTCACAGATTCACCTGCACTTTTCAACCACTGACTGG 3096  
QY 3114 CAGCACTCATCAACAATTTGGGATTCGGCCCAAGAGACTCAACTTCAACTCTTC 3173  
Db 3097 CAGCACTCATTAACAACAATTTGGGATTCGGCCCAAGAGACTCAACTTCAACTCTTC 3156  
QY 3174 AACATCCAAGTCAAGAGGTCAAGCAATGATGGCTCAACAACCATGCTAATACTT 3233  
Db 3157 AACATCCAAGTCAAGAGGTCAAGCAATGATGGCTCAACAACCATGCTAATACTT 3216  
QY 3234 ACCAGACGGTTCAGTCTTCTCGACTCGGAGTACCACTTCGTAAGTCTCGGCTCT 3293  
Db 3217 ACCAGACGGTTCAGTCTTACGCACTCGGAGTATCAGTCTCGGCTCGGCTCG 3276  
QY 3294 GCGCAACGAGGCTGCTCCCTCCGTTCCCGGAGCTGTCATGATTCGCAATACGGC 3353  
Db 3277 GCGCAACGAGGCTGCTCCCGGTTCCAGCGAGCTTCATGCTCCCTCAGTATGA 3336  
QY 3354 TACCTGACGCTCAACAATGGCAGCCAGCCGTTGAGAGCTTCCTTTACTGCTGGA 3413  
Db 3337 TACCTGACGCTCAACAATGGCAGCCAGCGGTGAGAGCTTCCTTTACTGCTGGA 3396  
QY 3414 TATTTCCCTTCTCAGATGCTGAGAACGGGCAACAACCTTACCTTCACTACCTTGA 3473  
Db 3397 TACTTCCCTTCCAGATGCTAAGACTGGAATACTTCAATTCAGCTATACCTTGA 3456  
QY 3474 GAAGTCCCTTCCACAGAGCTACCGGCAACAGCCAGAGCTTGACCGGCTGATGA 3533  
Db 3457 GATGTACCTTTTCAACAGAGCTACGCTCAACAGAGCTTGATGCTTGATGA 3516  
QY 3534 CTGATGACCAATACCTGTATTAATCTGAACAGAACTCA--AAATCAGTCCGGAAGTGC 3590  
Db 3517 CTATTTGATCAGTATCTGTACTACTGAACAGAAAGCAAGCAAACTCTGGAACAA 3576  
QY 3591 CAAAAAAGGACTGTGTTTGAAGCGGTGGGTCTCCAGCTGCACTGCTTCAAGCCCAA 3650  
Db 3577 AACCAATCAAGGCTGCTTTTGAAGCGGTGGGTCTCCAGCTGCACTGCTTCAAG 3636  
QY 3651 AACTGGCTACCTGAGACCTGTATTCGGCAGCAGCGGCTTTCTAAACAAACAGACAA 3710  
Db 3637 AATGGCTACCTGAGACCTGTATTCGGCAGCAGCGGCTTTCTAAACAAACAGACAA 3696  
QY 3711 AACACACGCAATTTTACCTGAGCTGTGCTTCAAAATATACTCAATGGGCTGAATCC 3770  
Db 3697 AACACACGCAATTTTACCTGAGCTGTGCTTCAAAATATACTCAATGGGCTGAATCC 3756  
QY 3771 ATCATCAACCTTGCACTGCTATGCTCTCAACAAAGACGAGCAAGTCTTTCC 3830  
Db 3757 CTGCTGAATCCAGGACCACTATGCTCAGTCAAGGACGATGAAGAAAATTTTCCCT 3816  
QY 3831 ATGAGCGGTGTCATGATTTTGAAGAAAGAGAGCGCGGAGCTTCAAACTGCAATGGAC 3890  
Db 3817 ATGACGCGCAATCTAATATTTGCAAGAAAGGAGCAACGCGCAAGTAAAGCAATGAT 3876  
QY 3891 AATGTCATGATTAACAGACGAAGGAATTAAGCCACTAATCCTGTGGCCACCGAAGA 3950  
Db 3877 AATGTAATGATTAACAGACGAAGGAATTAAGCCACTAATCCTGTGGCCACCGAAGA 3936  
QY 3951 TTTGGGACCGTGGCAGTCAATTTCCAGAGCAGCAGACAGCCCTGCGACCGAGATG 4010  
Db 3937 TATGGAATCTGTGCAATACTTGCAGAGCTCAATATACGCTCCACGACTGGAATGTC 3996

QY 4011 CATGCTATGGAGCAATTAACCTGGCATGTGTGGCAAGATAGACGCTGTACCTGAGGCT 4070  
Db 3997 AATCATCAGGGGCTTACCTGGCATGTGTGGCAAGATAGACGCTGTACCTGAGGCT 4056  
QY 4071 CCCATTTGGGCCAAATCTCTCAACAGATGACACTTCAACCCCTCTCTCTTAAGGC 4130  
Db 4057 CCTATCTGGGCAAGATCTCTCAACAGATGACACTTCACTCTCTCTCTGATGGGA 4116  
QY 4131 GGCCTTGGACTCAAGAACCCGCTCTCAGATCTCTCAATAAAACACGCGCTGCTGCG 4190  
Db 4117 GGCCTTGGACTCAAGAACCTCCCTCTCAATAATCATGATCAAAATACTCCGCTACCGCA 4176  
QY 4191 AATCTCCGCGGAGTTTCAAGCTACAAAGTTTGTCTTCAATCAACCAATACTCACA 4250  
Db 4177 AATCTCCGAGCACTTCAAGCCCGCAAGTTTGTCTTCAATCACTCAGTACTCACT 4236  
QY 4251 GACCAAGTGAAGTGAAGTGAATGGAGCTGCAAGAAAGAAACAGCAAGCGCTGGAAT 4310  
Db 4237 GACCAAGTGAAGTGAAGTGAATGGAGCTGCAAGAAAGAAACAGCAAGCGCTGGAAT 4296  
QY 4311 CCCGAAGTGAAGTGAAGTGAATGGAGCTGCAAGAAAGAAACAGCAAGCGCTGGAAT 4370  
Db 4297 CCAGAGATTAAGTGAAGTGAATGGAGCTGCAAGAAAGAAACAGCAAGCGCTGGAAT 4356  
QY 4371 AACATGAGCTTATATGAGCTGCGCCCATTTGGCAACCCGTACTACCCGCTGCTG 4430  
Db 4357 ACTAATGAGTGTATATGAGCTGCGCCCATTTGGCAACCCGTACTACCCGCTGCTG 4416  
QY 4431 TAAATGAGTGTATATGAGCTGCGCCCATTTGGCAACCCGTACTACCCGCTGCTG 4486  
Db 4417 TGAATCTGTTATATGAGCTGCGCCCATTTGGCAACCCGTACTACCCGCTGCTG 4476  
QY 4487 TGTCTTCTTATCTTATGAGCTGCGCCCATTTGGCAACCCGTACTACCCGCTGCTG 4540  
Db 4477 ACTTCTTATCTTATGAGCTGCGCCCATTTGGCAACCCGTACTACCCGCTGCTG 4536  
QY 4541 GTTCCGCTTCCGATTAAGAC-----TTACGTCATCGGCTTACCCCTAGTGA 4588  
Db 4537 CTTCGCTTCCGATTAAGAC-----TTACGTCATCGGCTTACCCCTAGTGA 4596  
QY 4589 TGAAGTTGCCCACTCCCTCTCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4648  
Db 4597 TGAAGTTGCCCACTCCCTCTCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4656  
QY 4649 TCCGCAAGCGGAGAGCTCTCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4708  
Db 4657 TCCGCAAGCGGAGAGCTCTCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4716  
QY 4709 GAGTGGCAA 4718  
Db 4717 GAGTGGCAA 4726

RESULT 4  
AF028705 4722 bp DNA linear VRL 12-JAN-1998  
LOCUS AF028705  
DEFINITION Adeno-associated virus 3B, complete genome.  
ACCESSION AF028705  
VERSION AF028705.1 GI:2766608  
KEYWORDS  
SOURCE adeno-associated virus 3B.  
ORGANISM adeno-associated virus 3B.  
REFERENCE 1 (bases 1 to 4722)  
AUTHORS Rutledge, E.A., Halbert, C.L. and Russell, D.W.  
TITLE Infectious clones and vectors derived from adeno-associated virus (AAV) serotypes other than AAV type 2  
JOURNAL J. Virol. 72 (1), 309-319 (1998)  
MEDLINE 98080418  
PUBMED 9420229  
REFERENCE 2 (bases 1 to 4722)  
AUTHORS Rutledge, E.A. and Russell, D.W.  
TITLE Direct Submission



JOURNAL Submitted (25-SEP-1997) Hematology, Univ of Washington, Box 357720, Seattle, WA 98195, USA  
FEATURES Location/Qualifiers  
source 1..4722  
    /organism="adeno-associated virus 3B"  
    /db\_xref="taxon:68742"  
    /note="AAV3B"  
CDS  
    317..2191  
        /codon\_start=1  
        /product="nonstructural protein Rep78"  
        /protein\_id="AAB95451.1"  
        /db\_xref="GI:2766609"  
        /translation="MPGFYEVLKVPSPDLDEHLPGISNPFVNWVAEKEMELPPDSMD  
PNLIEQAPLTVAEKLOREPIYEMRVRVSKAPEALFEVQEKETYPHLHLIETIGVKS  
MVGKRVISOIKEKLVTRIRGVPEPOLPNWPAVTKIRNGAGGNKVDDCYIPNVLPRK  
TOBELQWMTNMDQYLSACINLAEKRLVAOHLTEVSOTQENKKNONPNSDAVRS  
KTSARYMELVGMVLDRGITSEKOWIQEDQASYISENANSRSQIKALDNASKIMSL  
TKTAPDYLGVSNPEDITKRIYQILENGYDPOYASVFLGWAOKKFGKRTIMLFG  
PATGKTNIABEIAIAHAVPYGCVNWTNENPFNDQVDMVIMWEEKMTAKVESAKA  
ILGSKVRVDQCKSSAQIEPTPIVTSNTNMCVIDGNSTFEHQPLQDRMEFEEL  
TRRLDHDGKVTQKQEVKDFEWAADHVTDAHEFYRVGAKKRPASNDADVSPPKQ  
CTSLAQPTTSDAEAPADYADRYQNKCSRHVGMNMLFPCKTCERMOISNVCTHGOR  
DGECEPFGMSSESQPVSVKKTYQKLCPIHHLIFAEPIACSACDLANVDLDCVSEQ"

2208..4418  
    /codon\_start=1  
    /product="capsid protein VP1"  
    /protein\_id="AAB95452.1"  
    /db\_xref="GI:2766610"  
    /translation="MAADGYLPDWMLEDNLSEGIREWMAKPGVPOPKANOQHODNRG  
LVLPGYKYLGPNGLDKGEVNEADAALAEHDKAYDQOLKAGDNPYLKYNHDAEFQE  
RLQEDTSFGNIGRAVFOAKRILEPLGLVEBAKTAPEGKRPVDSPOEPSSSGVG  
KSGKQAPARKRLNFGQTDSEVPDPQPLGEPAPATSLGSNTMASGGAPMADNNEGA  
DVGNSGNMHCDSQWLGDRIYITSTRWALPTYNMHLKQISSQASNDNHFYGS  
TPWGYFDENRFHCHFSPRDWORLINNMWFRPKLSFKLNIQVKEVTQNDGTTIAN  
NLSTVOVFTDSEYOLPYVIGSAHQGLPPFPADVFMVPOQYGLTLNNGSOAVGRSSE  
YCLEYFPQSOMLRGTNNFQSYTFEDVPHSSYAHQSILDRIMNPLIDQYLYLNRTQG  
TTSGTNOSRLIFSQAGPQMSLQARNMLPGPCYQORLSKTANDNNSNFPMTAASK  
YHLNGRDSLVPNGPAMASHDDEEKFPPMHGNLIFKEGRTASNAELDNVMTDEEEL  
RTNPNVATEQYGTVANNLQSNATAPTTRTVNDGALPDMWQDRDVLQGPIMAKIPH  
TDGFHPSPLMGFGFLKHPQIMIKNTVPANPPTTFSPAKFASFITQYSTGVASVE  
IEMELQENSKRMNPEIQYTSNKNKSVNVDFTVDJNGVYSEPRPIGTRYLTRNL"

BASE COUNT 1204 a 1297 c 1243 g 978 t  
ORIGIN

Query Match 67.9%; Score 3205; DB 14; Length 4722;  
Best Local Similarity 81.6%; Pred. No. 0;  
Matches 3872; Conservative 0; Mismatches 820; Indels 55; Gaps 12;

QY 2 TGCCCACTCCCTCTCTGCGCGCTCGCTCGCTCGGTGGGCGCTGCGCAACCAAGTCCGCA 61  
DB 1 TGGCACTCCCTCTATGCGCACTCGCTCGCTCGGTGGGCGCTGCGCAACCAAGTCCGCA 60  
QY 62 GACGCGAGAGCTCTGCTCTGCGCGCGTCCGCAACGAGCGAGCGCGCAGAGAGGAGTGG 121  
DB 61 GACGCGAGTGTCTTGACAGTCCGCCACCGAGCGAGCGAGTGCATAGAGGAGTGG 120  
QY 122 GCAACTCCATCACTAGGGGTAATCGGAAAGCGCTCCACGCTGCCGCGTCAAGCGTGA 181  
DB 121 CCAACTCCATCACTAGAGGTATGCGAGTGAAGCGGCGGCGGCGGAGGAGACAC 180  
QY 182 GTAAATTAAGTATAGGGAGTGTCTGTATTAAGTGTCACTGAGTGTCTTTGGACA 241  
DB 181 GCCTACCA-----GCTGCGTCAAGCACTCAAGTGA-CCCTTTGGACA 222  
QY 242 TTTTGGACACCAACGTCGTCATTGAGGTATATATGCGCGAGTGAAGCG-AGCAGATCTC 300  
DB 223 GTTGGACACCAACGTCGTCGAGGTATATATTCTGAGTGAAGCGAACGAGAGCTC 282  
QY 301 CATTTGACCGCGAAATTTGAACGAGCAGCAGCCATGCGGCGCTTCTACGAGATCGTAT 360  
DB 283 CATTTGACCGCGAAATTTGAACGAGCAGCAGCCATGCGGCGCTTCTACGAGATGTCTCT 342

QY 361 CAAGTGCAGCGACCTGACGACCACTGCGCGGCAATTCTGACTCGTTGTGAGCTG 420  
DB 343 GAAGTCCCGAGTGACCTGACGACCACTGCGCGGCAATTCTGACTCGTTGTGACTG 402  
QY 421 GGTGCGCGAGAGAAATGGAGCTGCGCGGATTTCTGATGATCTGAATCTGAATTGA 480  
DB 403 GGTGCGCGAGAGAAATGGAGCTGCGCGGATTTCTGATGATCTGAATCTGAATTGA 462  
QY 481 GCAAGCAACCCCTGACCGTGGCGGAGAGCTGACGCGGCACTTCTGTCGAATGGCGCG 540  
DB 463 GCAAGCAACCCCTGACCGTGGCGGAGAGAGCTTCAAGCGGAGTCTGAGTGGCGCG 522  
QY 541 CGTGAATAAGGCGCGGAGCGCTCTTCTTGTTCAGTTCCAGAGGGCGAGTCTTACTT 600  
DB 523 CGTGAATAAGGCGCGGAGCGCTCTTCTTGTTCAGTTCCAGTTCCAGAGGGCGAGTCTT 582  
QY 601 CCACCTCCATATTCTGTTGAGACCAACGCGGGTCAATTCATGTGTCTGGCGCTTCT 660  
DB 583 CCACCTGACGCTGATTGAGACCATCGGGGTCAATTCATGTGTGTGGCGCTTACGT 642  
QY 661 GAGTCAGATTAGGACCAAGCTGTGACAGACCATCTACCGCGGATCGAGCCGACCTGCG 720  
DB 643 GAGTCAGATTAAAGAGAGCTGTGACCGCATCTACCGCGGGTCAAGTCGAGCCGACGTTCC 702  
QY 721 CAATGCTTCCGCTGACCAAGACGCGTAATGCGCGGAGGGGGAACAAGTGTGGA 780  
DB 703 GAATGCTTCCGCTGACCAAGACGCGTAATGCGCGGAGGGGGAACAAGTGTGGA 762  
QY 781 CGAGTCTACATCCCACTACCTCTGCGCAAGACTCAGCCCGAGCTGACGTGGCGGTG 840  
DB 763 CGAGTCTACATCCCACTACCTCTGCGCAAGACTCAGCCCGAGCTGACGTGGCGGTG 822  
QY 841 GACTAACATGAGGAGTATATAAGCGCTGTTGAACCTGCGCGGAGCGCAACCGCTCGT 900  
DB 823 GACTAACATGAGGAGTATATAAGCGCTGTTGAACCTGCGCGGAGCGTAACCGCTCGT 882  
QY 901 GCGGCAAGCACTGACCCGAGTCAAGCAGACCCAGAGCAGAGCAACAAGAGAACTGAACCC 960  
DB 883 GCGGCAAGCACTGACCCGAGTGTGAGAGCAGAGCAGAGCAACAAGAGAACTGAACCC 942  
QY 961 CAATTCGACCGCGCTGTCACTCGGTCAAAAACCTCGCGGCTACATGAGCTGTGCG 1020  
DB 943 CAATTCGACCGCGCTGTCACTCGGTCAAAAACCTCAAGGATACATGAGCTGTGCG 1002  
QY 1021 GTGCTGTGACCGCGGATCACTCCGAGAGAGTGAATCCAGAGACCAAGCGCTC 1080  
DB 1003 GTGCTGTGACCGCGGATCACTCCGAGAGAGTGAATCCAGAGACCAAGCGCTC 1062  
QY 1081 GTACATCTCTTCAACGCGCTTCCAACTCGCGGTCCAGATCAAGCGCTTGAACA 1140  
DB 1063 GTACATCTCTTCAACGCGCTTCCAACTCGCGGTCCAGATCAAGCGCTTGAACA 1122  
QY 1141 TGCCGCAAGATCATGCGCTGACCAAAATCCGCGCGCACTACCTGTAGCGCGCTCC 1200  
DB 1123 TGCTTCAAGATCATGAGCTGACAAAGACGGCTCCGACTACCTGTGCGCAGCAACC 1182  
QY 1201 GCGCGGACATTAACCAACCAACGATCTACCGCATCTGAGCTGAACGCGCTACGAACC 1260  
DB 1183 GCGCGGACATTAACCAAAATCGATCTACCAATCTTGAAGCTGAACGCGCTACGAATCC 1242  
QY 1261 TGCTACGCGCGCTCGCTTCTCTGCGGTGGGCCAGAAAAGTTGCGGAAGCGCAAC 1320  
DB 1243 GCAATACGCGCGCTCGCTTCTCTGCGGTGGGCCAGAAAAGTTGCGGAAGCGCAAC 1302  
QY 1321 CATCTGCTGTTGGCGGCGCAACGAGGCAAGCAATCGCGGAAGCATCGCCCA 1380  
DB 1303 CATCTGCTGTTGGCGGCGCAACGAGGTAACCAACATCGCGGAAGCATCGCCCA 1362  
QY 1381 CGCGGTGCGCTTCTACGCGTGGCTCACTGACCAATGAGAACTTCCCTTCAATGATG 1440  
DB 1363 CGCGGTGCGCTTCTACGCGTGGCTTAACTGACCAATGAGAACTTCCCTTCAACGATG 1422  
QY 1441 CGTCGACAAAGATGTGATCTGTTGGAGGAGGCGCAAGATGACGCGCAAGGTGCTGAGTC 1500

Db	1423	CGTCGACAGATGGTGAATCTGGTGGGAGGAGGGCAAGATGACGGCCAAAGTCTGTGAGAG	1482
Qy	1501	CGCCAAAGGCATTCTCGGCGGCAGCAAGGTGCGCGTGACCAAAAGTGCAAGTCGTCGC	1560
Db	1483	CGCCAAAGGCATTCTGGGCGGAAGCAAGGTGCGCGTGACCAAAAGTGCAAGTCATCGGC	1542
Qy	1561	CCAGATGACCCCCAACCCCGTGATCGTCACTTCCAACCAACATGTGCGCGGTGATTGA	1620
Db	1543	CCAGATGAAACCACTCCCGTGATCGTCACTTCCAACCAACATGTGCGCGGTGATTGA	1602
Qy	1621	CGGGAACAGCACCACTTCGAGCACAGCAGCCGTTGCAAGACCGGATGTTCAATTGA	1680
Db	1603	CGGGAACAGCACCACTTCGAGCATCAGCAGCCGTTGCAAGACCGGATGTTAAATTGA	1662
Qy	1681	ACTCACCCGCGCTTGAGCATGACTTTGGCAAGTGACAAAGCAGGAAGTCAAGACTT	1740
Db	1663	ACTTACCCGCGCTTTGACCATGACTTTGGGAAGGTCAACAAACAGGAAGTAAAGTACTT	1722
Qy	1741	CTTCGCTGGGCGCAGGATCAGTGACCGAGTGCGCATGAGTTCTACGTCAAGAAAGG	1800
Db	1723	TTTCGGTGCGCTTCGATCAGTGACTGACGTGGCTCATGAGTTCTACGTCAAGAAAGG	1782
Qy	1801	TGGAGCCAACAAAGAGCCGCCCCGATGACGCGGATAAAAGCAGCCCAAGCGGCGCTG	1860
Db	1783	TGGAGCTAAGAAACGCCCCGCTTCAATGACGCGGATGTAAGCAGCCAAACGCGCAGTG	1842
Qy	1861	CCCCCTAGTCGCGGATCCATCGACGTCAAGCGCGGAAGAGCTCCGGTGACTTTGCCGA	1920
Db	1843	CACGTCACTTGGCGAGCCGCAACAGTCAAGCGCGGAAGAGCTCCGGTGACTTTGCCGA	1899
Qy	1921	CAGGTACCAAAACAATGTTCTCGTCAAGCGGCGCATGCTTCAGATGCTGTTCCCTGCA	1980
Db	1900	CAGGTACCAAAACAATGTTCTCGTCAAGCGGCGCATGATCTGATGCTTTTCCCTGTA	1959
Qy	1981	GACATGCGAGAGAATGAATCAGAAATTTCAACATTTGCTTCAAGCAGCAGGAGAGAGACTG	2040
Db	1960	AACATGCGAGAGAATGAATCAAAATTTCCAATGTCTGTTTAAAGCATGTCAAAAGAGACTG	2019
Qy	2041	TTCAGAGTGTCTCCCGGCGGTGTCAGAAATCTCAAC-----CGTCTCAGAAAGAGAC	2094
Db	2020	TGGGGAATGCTTCCCTGGAATGTCAGAATCTCAACCCGTTTCTGTCTCAAAAAGAGAGAC	2079
Qy	2095	GTATCGGAAACTCTGTGCCATTCATCATCTGTGCGGCGGCTCCGAGATTGCTTCTC	2154
Db	2080	TTATCAGAAACTGTGTCCAATTCATCATATCTGTGGAAGGCAACCGAGATTGCTTCTC	2139
Qy	2155	GGCCTCGATCTGGTCAACGTGACCTTGATGACTGTGTTCTGAGCAATAAATGACTTA	2214
Db	2140	GGCCTCGATTTGGCCATGTGACCTTGATGACTGTGTTCTGAGCAATAAATGACTTA	2199
Qy	2215	AACCAAGTATGCTGCCGATGTATCTTCCAGATTGGCTCGAGGACAACCTCTTGAGG	2274
Db	2200	AACCAAGTATGCTGCTGACGSETTATCTTCCAGATTGGCTCGAGGACAACCTTCTGAGG	2259
Qy	2275	GCAATTCGCGAGTGTGGGACTTGAAACCTTGAGCCCGGAAGCCAAAGCCAAACGACAA	2334
Db	2260	GCAATTCGAGTGTGGGCTTGAAACCTTGAGTCCCTCAACCCAAAGCCAAACGACAA	2319
Qy	2335	AGCAGGACGACGCGCGGCTGTGTGCTTCTGAGTACAAAGTACCTCGAACCTTCAAG	2394
Db	2320	ACCAGGACAAACCGTGGGGTCTGTGCTTCCGGGTTACAAATACCTCGGAACCGGTTAAG	2379
Qy	2395	GACTCGACAAAGGGAGCCCGTCAACCGCGCGGACGCGAGCGGCTCGAGCAGCAGAG	2454
Db	2380	GACTCGACAAAGGAGAGCCGCTCAACGAGGCGGACGCGGAGCCCTCGAACACGACAAAG	2439
Qy	2455	CCTACGACCAAGCAGCTCAAAAGCGGTGACAATCCGTACCTGCGGTATAACCAAGCCGACG	2514
Db	2440	CTTACGACCAAGCAGCTCAAGGCGGTGACAACCCGTACCTCAAGTACCAACCAAGCCGACG	2499
Qy	2515	CCGAGTTTCAGAGCGTCTGCAAGAAATACGTCTTTTGGGGCAACCTCGGCGAGAG	2574

D	b	2500	CCGAGTTTCAGGAGCGCTCTTCAAGAAAGATACGTCTTTTGGGGGCAACCTTGGCAGAGCAG	2559
O	y	2575	TCTTCCAGGCCAAGAACGGGTTCTGAACTCTCGGTCTGGTTGAGGAAGCGCTAAGA	2634
D	b	2560	TCTTCCAGGCCAAGAAAGAGATCCTGAGCCTCTTGCTGGTTGAGGAAGCAGCTAAAA	2619
O	y	2635	CGGCTCCTGGAAAAAGAAAGTCCGGTAGAGCAGTGCACAAGACCAGACTCCTCCGG	2694
D	b	2620	CGGCTCCTGGAAGAAGAGGCGCTGTAGATCAGTCTCCTCAGGAACCGACTCATATCTG	2679
O	y	2695	GCATCGGCAAGACAGGCCAGCAGCCCCGCTAAAAAGAGACTCAATTTTGGTCAGACTGGCG	2754
D	b	2680	GTGTTGGCAAAATCGGGCAAAACAGCCTGCCAGAAAAAGACTTAATTTCCGTCAGACTGGCG	2739
O	y	2755	ACTCAGAGTCAGTCCCCGATCCAACTCTCGAGAACTTCAGCAACCCCGCTGCTG	2814
D	b	2740	ACTCAGAGTCAGTCCAGACCCCTCACTCTCGAGAACCAACAGACGCCCCCAAGTT	2799
O	y	2815	TGGAGCTTACTACAATGCTTCAGGCGGTGGCGCACCAATGGCAGACAATAAGAAAGCG	2874
D	b	2800	TGGATCTATATACATAGGCTTCAGGCGGTGGCGCACCAATGGCAGACAATAAGAGGTG	2859
O	y	2875	CCGACGAGTGGGTAATGCTCAGAAATTGGCATTTGCCATTCCATAGGCTGGCGACA	2934
D	b	2860	CCGATGAGTGGGTAATTCCTCAGAAATTGGCATTTGCCATTCCAAATGCTGGCGACA	2919
O	y	2935	GAGTCATCAACCAACGACCCCGACTGGGCTTGCCCACTACATAACCACTCTACA	2994
D	b	2920	GAGTCATCAACCAACGACCAAGACTGGGCGCTGCCACTTACAACAACCATCTTACA	2979
O	y	2995	AGCAAAATCTCAGTGCTTCAACGGGGGCCAGCAACGACAACCACTACTTCGGCTACAGCA	3054
D	b	2980	AGCAAAATCTCAGGCCAATCA--GAGCTTCAAAACGACAACCACTACTTTGGCTACAGCA	3036
O	y	3055	CCCCCTGGGGTATTTTGATTTCAACAGATTCCACTGCCATTTTCAACCACTGACTGGC	3114
D	b	3037	CCCCCTGGGGTATTTTGACTTTAACAGATTCCACTGCCATTTCAACCACTGACTGGC	3096
O	y	3115	AGGACTCATCAACAACAATTGGGATTCGGCCCAAGAGACTCAACTTCAAACTCTTCA	3174
D	b	3097	AGGACTCATTAACAACAATGGGATTCGGCCCAAGAACTCAGCTTCAAGCTTCA	3156
O	y	3175	ACATCCAAGTCAAGAGTCAACGACGAATGATGCGCTCAACAACCATCGTAATACTTA	3234
D	b	3157	ACATCCAAGTAAAGAGTCAACGACGAACGATGGCAACGACTATTGCCAATAACTTA	3216
O	y	3235	CCAGCAGGTTCAAGTCTTCTCGACTCGGAGTACCAGTTCCGTACGTCTCGGCTCG	3294
D	b	3217	CCAGCAGGTTCAAGTGTATTACGGACTCGGAGTATCAGTCCCGTACGTCTCGGCTCG	3276
O	y	3295	CGCACCAGGCTGCTCCTCCCTCCGTTCCCGCGGACGTGTATGATTTCCGCAATACGGCT	3354
D	b	3277	CGCACCAGGCTGCTCCTCCGCGTTCCAGCGGACTCTTATGTCCTCAGTATGAT	3336
O	y	3355	ACCTGACGCTCAACAATGCGACGCAAGCCGTGGGACGTTCATCTTTACTGCTGGAAT	3414
D	b	3337	ACCTGACGCTGAACAACGGAAGTCAAGCGGTGGGACGCTCATCTTTTACTGCTGGAAT	3396
O	y	3415	ATTTCCCTTTCAGATGCTGAGAACGGGCAACAATTACCTTCACTACACTTGTAGAG	3474
D	b	3397	ACTTCCCTTTCAGATGCTAAGGACTGGAATTAATTCAATTCACTATATACCTTGTAGAG	3456
O	y	3475	AAGTGCCTTTCACAGCAGCTACGGGCAACGCCAGAGCCGTGGAACGGCTGATGATCTC	3534
D	b	3457	ATGTACCTTTCACAGCAGCTACGCTCACAGCCAGAGTTTGATCGCTTGATGATCTC	3516
O	y	3535	TCATGACCAATACCTGTATTACCTGAACAGACTCA--AAATCAGTCCGGAAGTGGCC	3591
D	b	3517	TTATTGATCAATATCTGTACTACTGTAACAGAACGACGACAGAAACAACCTCTGGAACAACA	3576
O	y	3592	AAAAACAAGACTTGTCTGTTTAGCCGTGGGTCTCCAGCTGGCATGTCTGTTACGCCCAAAA	3651
D	b	3577	ACCAATCAGGCTGCTTTTATGACGAGGCTGGGCTCAGTCTATGTCTTTGACGAGCCAGAA	3636



QY	3652	ACTG	GC	TAC	CTG	GAC	CC	TG	TAT	CGG	CAG	CAG	CG	CG	TTT	CTA	AAA	CA	AAAA	CA	GAC	A	CA	3711
Db	3637	ATTG	GC	TAC	CTG	GGC	CC	TG	CTA	CCG	CA	CAG	AG	ACT	TTT	CA	A	AG	ACT	GT	CTA	AC	GA	3696
QY	3712	ACA	CAG	CA	AT	TTT	TAC	CTG	GAC	TG	TG	CTT	CA	AAA	AT	TA	A	CT	CA	TG	GG	CG	TGA	3771
Db	3697	ACA	CAG	TA	ACT	TTT	CC	T	TG	GAC	AG	CG	CC	AG	CA	AA	TAT	CA	T	CA	T	AG	CG	3756
QY	3772	TCAT	CA	ACC	CTG	GCA	CTG	TAT	AG	CC	CT	CA	CA	AA	GA	CGA	GA	GA	CA	AG	TT	CTT	CC	3831
Db	3757	TG	GT	GA	AT	CC	AG	GA	CC	AG	CT	AT	AG	CC	AG	TC	CA	AG	AG	CA	GA	AA	AT	3816
QY	3832	TG	AG	CG	GT	GC	AT	GAT	TTT	TG	GA	AA	AG	AG	AG	CG	CC	CG	GA	CTT	CA	AA	CA	3891
Db	3817	TG	CA	CG	CA	AT	CTA	TAT	TTT	TG	GC	AA	GA	GG	GA	CA	CG	CA	AG	TA	AC	GA	AT	3876
QY	3892	AT	GC	TAT	GAT	TAC	AG	CGA	AG	AG	AA	ATT	AA	AG	CC	ACT	AA	CC	CT	GT	GG	CC	3951	
Db	3877	AT	GT	AT	GAT	TAC	CGAT	GAG	AA	GA	AG	ATT	CG	TAC	CA	CA	AT	CT	GT	GG	CA	AG	3936	
QY	3952	TTG	GAC	CG	TG	CAG	TCA	ATT	TC	CAG	AG	CAG	CAG	CA	CA	CC	CTG	CA	CC	CG	AG	AT	4011	
Db	3937	AT	G	AA	CT	GTG	CA	AA	TAC	TG	CA	AG	CT	CA	AA	TAC	AG	CT	CC	AG	CT	AG	3996	
QY	4012	AT	G	CT	AT	TG	GAG	CA	TT	AC	TG	GC	AT	G	TG	GC	AA	GA	TAG	AG	AC	GT	4071	

OY		4712 TGGGCAA 4718 	
Db		4716 TGCCCAA 4722	
RESULT 5			
LOCUS	AX282480	4679 bp	DNA linear PAT 02-NOV-2001
DEFINITION	Sequence 1 from Patent WO016888.		
ACCESSION	AX282480		
VERSION	AX282480.1 GI:16609611		
KEYWORDS			
SOURCE	adeno-associated virus 2. adeno-associated virus 2.		
ORGANISM	Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.		
REFERENCE	1 Xiao,W. and During,M.J. Production of chimeric capsid vectors Patent: WO 016888-A 1 20-SEP-2001; Neurologix, Inc. (US)		
AUTHORS			
TITLE			
JOURNAL			
FEATURES	Location/Qualifiers		
source	1..4679 /organism="adeno-associated virus 2" /db_xref="taxon:10804"		
BASE COUNT	1198 a 1262 c 1255 g 964 t		
ORIGIN			
Query Match	65.9%; Score 3107.8;	DB 6;	Length 4679;
Best Local Similarity	80.2%;	Pred. No. 0;	
Matches 3790;	Conservative 0;	Mismatches 882;	Indels 53;
Gaps			10;
OY	1 TTGCCACTCCCCTCTGTGCGCGCTCGTGCCTCGTGCGGAGCCTCGGACCMAAGTCCGC	60	
Db			
OY	1 TTGGCCACTCCCCCTCTGTGCGCGCTCGTGCCTCGTCACTGAGGCCGGGCGAACAAAGTCGCC	60	
Db			
OY	61 AGACGGCAGAGCTCTGCTCTGCGCGCGCCCGCCACCGAGCGAGCGCGCAGAGAGGAGTG	120	
Db			
OY	61 CGACGCCCGGCGTTTGCCCGCGCGCCTCAGTAGCGAGCGAGCGCGCAGAGAGGAGTG	120	
Db			
OY	121 GGCACTCCATCACTAGGGGTAATCGGAAGCGCCTCCGACGCTGCGCGCTGAGCGCTGA	180	
Db			
OY	121 GCCAAGTCCATCATCATTAGGGGT-----TCCTGAGGGGTGAGTCTGA	163	
Db			
OY	181 CGTAATTACGTCATAGG---GAGTGTCTGTATTAGCTGCACGTGAGTGTCTTTGC	237	
Db			
OY	164 CGTGAATTACGTCATAGGTTAGGAGGTCCTGTATTAGAGTCACGTGAGTG-TTTTGC	222	
Db			
OY	238 GACATTTTGGCACACCACTGGGCCATTAGGTTATATATGSCCGAGTGAGCGAGCAGAT	297	
Db			
OY	223 GACATTTTGGCACACCACTGTGTCAAGCTGGGTATTAAAGCCCGAGTGAGCAGCAGGGT	282	
Db			
OY	298 CTCATTTTGG-ACCGCGAAATTGAACGAGCAGCAGCAGCCATGCGCGGCTTCTACGAGATCG	356	
Db			
OY	283 CTCCATTTTGAAGCGGAGGTTGAACGCGCAGCGCCATGCGCGGGTTTTACGAGATTG	342	
Db			
OY	357 TGATCAAGGTGCCGAGCGACCTGAGCAGCAGCACCCTGCCGGGCACTTCTGACTGTTGTGA	416	
Db			
OY	343 TGAATAAGGTCCCCAGCGACTTGACGAGCATCTGCCCGCAATTCTGACAGCTTGTGA	402	
Db			
OY	417 GCTGGGTGCCGAGAAGAAATGGAGCTGCCCCCGGATTTGACATGATCTGAATCTGA	476	
Db			
OY	403 ACTGGGTGCCGAGAAGAAATGGAGTTGCCCGCAGATTCTGACATGATCTGAATCTGA	462	
Db			
OY	477 TTGAGCAGGCAACCCCTGACCGTGCCCGAGAAGCTGCAGCGCGACTTCTGTGTCGAATGCG	536	
Db			
OY	463 TTGAGCAGGCAACCCCTGACCGTGCCCGAGAAGCTGCAGCGCGACTTCTGACGAATGCG	522	
Db			
OY	537 GCCCGGTGAGTAAGCCCCCGAGAGCCCTCTTTGTTGTTGATTCAGTTGAGAAAGGCGAGTCT	596	
Db			
OY	523 GCCGTGTGAGTAAGCCCCCGAGAGCCCTTTTCTTTGTGCAATTGAGAAGGAGAGAGCT	582	
Db			
OY	597 AATTCCACCTCATATTTCTGTGTGAGACAACGGGGGTCAATTCATGCTGTGGCGGCT	656	
Db			



Db 583 ACTTCCACATGCAAGTCTCTGTAAGAACACCGGGGTGAATCCATGTTTGGACGTT 642  
QY 657 TCCTGAGTCAGATTAGGACAAAGCTGTGTGAGACCATCTACCGCGGATCGAGCCGACC 716  
Db 643 TCCTGAGTCAGATTGCGGAAAAATGATTGAGAAATTTACCGCGGATCGAGCCGACTT 702  
QY 717 TGCCCAACTGTTTCGCGGTGACCAAGACGGTAAATGCGCGGAGGGGGAACAAGTGG 776  
Db 703 TGCCAAACTGTTTCGCGGTGACCAAGACCAAGAAATGCGCGGAGGGGGAACAAGTGG 762  
QY 777 TGACAGAGTGTACATCCCCCACTACCTCTGCCCCAAGACTCAGCCCGAGCTGCAGTGG 836  
Db 763 TGATGAGTGTACATCCCCCACTACCTCTGCCCCAAGACTCAGCTGCAGTGG 822  
QY 837 CGTGAATACATGAGAGAGATATATAGCGCTGTTGAACCTGCGCGAGCGCAACGCGC 896  
Db 823 CGTGAATATATGGAACAGATTTAAGCGCTGTTGAATCTCAGCGAGCGTAAACGGT 882  
QY 897 TCGTGGCGAGCACTGACCCACGTACGCGAGACCCAGAGAGAGAAAGAAATCTGA 956  
Db 883 TGTGGCGAGCATCTGACGACGCTGTGCGAGACGAGAGAGAGAAAGAAATCAGA 942  
QY 957 ACCCAATTCAGACGCGCTGTCTATCCGGTCAAAAACTCCGCGGCTACATGAGCTGG 1016  
Db 943 ATCCCAATTCAGATGCGCGGTGATCAGATCAAAACTTCAGCCAGTACATGAGCTGG 1002  
QY 1017 TCGGCTGCTGTGTGAGACCGGGCATCACTCCGAGAGAGAGAGAGAGAGAGAG 1076  
Db 1003 TCGGCTGCTGTGTGAGACAGGGGATTACTTCGAGAGAGAGAGAGAGAGAGAG 1062  
QY 1077 CCTGATCATCTCTTCAACGCGCTTCCAACTCGCGGTCCAGATCAAGCGCTCTGG 1136  
Db 1063 CCTCATACATCTCTTCAATGCGGCTCCAACTCGCGGTCCAAATCAAGGCTGCTTGG 1122  
QY 1137 ACAATGCGGCAAGATCATGCGCTGACCAATCCGCGCGGACTACCTGTAGGCGCCG 1196  
Db 1123 ACAATGCGGCAAGATCATGAGCTGATCAAAACCGCGCGGACTACCTGTGCGGCGAGC 1182  
QY 1197 CTGCGCGCGGCAATTAAACCAACCGCATCTACCGCATCTGAGCTGAACGGCTACG 1256  
Db 1183 AGCCCGTGAAGACATTTCCAGCAATCGAATTTAATAATTTGGAATTAACGGGTACG 1242  
QY 1257 AACCTGCTACGCGCGCTCGCTTTCTGCGCTGCGGCGCAGAAAGTTGCGGAGCGCA 1316  
Db 1243 ATCCCAATATGCGGCTTCCGCTTTCTGCGGATGGGCGCAGAAAGTTGCGGAGAGGA 1302  
QY 1317 ACACCATGTGCTGTTTGGGCGCGCAACGCGGCAAGACCAACATCGCGGAGCCATCG 1376  
Db 1303 ACACCATGTGCTGTTTGGGCGCTGCAACTACCGGGAAGACCAACATCGCGGAGCCATAG 1362  
QY 1377 CCCACGCGGCTCTTCTACGCGCTGCTCAACTGGAACCAATGAGAACTTCCCTCAATG 1436  
Db 1363 CCCACATGTGCTCTTCTACGCGGTGCTAAACTGGAACCAATGAGAACTTCCCTCAACG 1422  
QY 1437 ATTGCTGCAAGAATGTGATCTGTGTGAGAGAGGGCAAGATGACGCGCAAGGTCTGG 1496  
Db 1423 ACTGTGTCGAAGAATGTGATCTGTGTGAGAGAGGGGAAGATGACGCGCAAGGTCTGG 1482  
QY 1497 AGTCCGCAAGGCCATTCTCGCGGAGCAAGGTGCGGTGACCAAAAGTGAAGTCTGT 1556  
Db 1483 AGTCCGCAAAAGCCATTCTCGAGAGAGCAAGGTGCGGTGACCAAAATGCAAGTCTCT 1542  
QY 1557 CCGCCAGATCGACCCCAACCCCGTGTGCTACCTCCAAACCAACATGTGCGCGCTGA 1616  
Db 1543 CCGCCAGATAGACCCGACTCCCGTGTGCTACCTCCAAACCAACATGTGCGCGCTGA 1602  
QY 1617 TTGACGGGAACAGCAACCACTTCGAGACCAAGCGGTTGACGAGCCGATGTTCAAT 1676  
Db 1603 TTGACGGGAACCTCAACGACCTTCGAACCAAGCAGCCGTTGCAAGACCGGATGTTCAAT 1662  
QY 1677 TTGAATCTACCCGCGCTGTGAGCATGACTTTGGCAAGGTGACAAAGCAAGAGTCAAG 1736  
Db 1663 TTGAATCTACCCGCGCTGTGATCATGACTTTGGGAAGGTCAACCAAGCAAGAGTCAAG 1722

QY 1737 AGTCTTCCGCTGGGCGCAGATCACTGACCGAGGTGGCCATGAGTTCTACGTACAGA 1796  
Db 1723 ACTTTTCCGCTGGGCAAGGATCACTGTTGAGGTGAGCATGAATTTCTACGTCAAAA 1782  
QY 1797 AGGTGAGACCAACAAAGACCCCGCCCGATGACCGCGGATAAAGCGAGCCCAAGCGGG 1856  
Db 1783 AGGTGAGACCAAGAAAGACCCCGCCCGATGACCGGATATAAGTGAGCCCAAGCGGG 1842  
QY 1857 CTGCCCCCTAGTGGCGGATCGACGTGACGCGGAGAGAGACTCCGTTGACTTTG 1916  
Db 1843 TGGCGAGTCACTTGGCAGCCATCGACGTGACGCGGAG--AGCTTCGATCAACTACG 1899  
QY 1917 CCGACAGTACCAAAACAAATGTTCTGTCACGCGGCGATGCTTCAAGATGTTCCCT 1976  
Db 1900 CAGACAGTACCAAAACAAATGTTCTGTCACGCGGCGATGATGATGTTCCCT 1959  
QY 1977 GCAAGCATGCGAGAGATGAATCAGATTTCAACATTTGCTTCAACGCGGAGAGAG 2036  
Db 1960 GCAGACAAATGCGAGAGATGAATCAGATTTCAATATCTGCTTCACTACGAGACAGAAAG 2019  
QY 2037 ACTGTTCAAGTGTCTCCCGCGGTGTCAAGATCTCAACCGGT--CGTCAGAAAGAGGA 2093  
Db 2020 ACTGTTAGAGTGTCTTCC--CGTGTCAAGATCTCAACCGGTCTGCTGTCAAAAGG 2076  
QY 2094 CGTATCGAAAATCTGTGCCATTCATCTGCTGCGGCGGCTCCGAGATGCTTGGT 2153  
Db 2077 CGTATCAGAAATCTGTCTACATTCATCATATCATGGAAGAGTGCCAGA--CGCTTGA 2133  
QY 2154 CCGCTGCGATCTGTCAACGTGGAAGCTGATGACGTGTCTTCTGAGCAATAAATGACTT 2213  
Db 2134 CTGCTGCGATCTGTCAATGTGAATTTGATGACGTGCACTTTGAAACATAAATGATTT 2193  
QY 2214 AAACAGGTATGCTGCCGATGTTATCTTCCAGATTTGGCTCAGAGACAACTCTGAG 2273  
Db 2194 AAATCAGGTATGCTGCCGATGTTATCTTCCAGATTTGGCTCAGAGACACTCTCTGAA 2253  
QY 2274 GGCATTCGCGAGTGTGGGACTTGAACCTGGAAGCCCGCAAGCCCAAGCCAGCA 2333  
Db 2254 GGAATAGACAGATGTGTGAAGCTCAAACTGCGCCACCAACCAAGCCCGCAGAGCGG 2313  
QY 2334 AAGCAGAGCAGCGCGCGGTGTGTCTTCTGCTTCAAGTAACTCTGAGACCTTCAAC 2393  
Db 2314 CATAGAGACAGCAGAGGGGTCTGTGCTTCTGCTTCAAGTAACTCTGAGACCTTCAAC 2373  
QY 2394 GGAATCAGCAAGGGGAGCCCGTCAACGCGGAGAGCGAGCGCCCTGAGCAGAGCAAG 2453  
Db 2374 GGAATCAGCAAGGGAGAGCGCGTCAACGAGCAGAGCGCGCGCCCTGAGCAGAGCAAA 2433  
QY 2454 GCCTACGACCAAGCATCAAGCGGCTGACATTCCTGCTGCTATACCAAGCGCGAC 2513  
Db 2434 GCCTACGACCGGAGCTGACAGCGGAGACAAACCGTACTCAAGTACAAACCGCGAC 2493  
QY 2514 GCCAGTTTCAAGAGCGGTCTGCAAGATACGTCTTTTGGGGGCAACCTCGGGCGAGCA 2573  
Db 2494 GCGGAGTTTCAAGAGCGGCTTTAAGAAATACGTCTTTTGGGGGCAACCTCGAGCGAGCA 2553  
QY 2574 GTCTTCAAGGCAAGAGCGGTTCTGAACCTCTGCTGCTGTTGAGGAAGGCGCTAAG 2633  
Db 2554 GTCTTCAAGGCAAGAGCGGTTCTGAACCTCTGCTGCTGTTGAGGAAGGCGCTAAG 2613  
QY 2634 ACGGCTCTGAAAAGAAAGTCCGCTAGAGCAGTCCGCAAGAGCCAGACTCTCTCG 2693  
Db 2614 ACGGCTCCGGGAAAAAGAGCGGCTAGAGCACTCTGTTGAGCCAGACTCTCTCG 2673  
QY 2694 GGCATCGGCAAGACAGGCGAGCGCGCTAAAGAAAGACTCAATTTTGTGACTGCGC 2753  
Db 2674 GGAACCGGAAAGCGGGCGAGCAGCTGCAAGAAAGAAATTTGATTTGTGACTGGA 2733  
QY 2754 GACTCAGAGTCACTCCCGATCCACAACCTCTCGAGAACTCCAGCAACCCCGCTGCT 2813  
Db 2734 GACGAGACTCACTACCTGACCCCGACCTCTCGGACAGCCACAGAGCCCTCTGCT 2793



PUBMED 7996133  
REFERENCE 2 (bases 1 to 4679)  
AUTHORS Berns,K.I., Bohenzky,R.A., Cassinotti,P., Colvin,D., Donahue,B.A.,  
Dull,T., Horer,M., Kleinschmidt,J.A., Rufling,M., Snyder,R.O.,  
Tratschin,J.-D. and Weitz,M.  
TITLE Direct Submission  
JOURNAL Submitted (15-JAN-1998) Cell Genesys Inc., 342 Lakeside Dr., Foster  
City, CA 94404, USA  
FEATURES  
source 1. .4679  
/organism="adeno-associated virus 2"  
/db\_xref="taxon:10804"  
/note="changes relative to the original sequence, GenBank  
Accession Number J01901, have been detected and verified  
by several different laboratories"  
repeat\_region 1. .145  
/note="inverted terminal repeat"  
/rpt\_type=inverted  
misc\_feature 42. .83  
/note="flip oriented DNA"  
precursor\_RNA 287. .4451  
/codon\_start=1  
/product="Rep 68 protein"  
/protein\_id="AAC03774.1"  
/db\_xref="GI:2906017"  
/translation="MPGFYEIVIKVPSDLDEHLPGISDSFVNVAEKEMELPPDSMD  
LNLIQAPLTVAEKIQDFLEMRVRVSKAPEALFVQFEKESYFHMVLVETGKXS  
MVLGRFLSQIREKLIQRIYRGIEPTLPNFAVTKTRNGAGGNKVDECYIPNYLLPK  
TOPELQAWTMEQYLACNLNTERKRLVAQHLTHVSQTOENKENONPNSDAPYTRS  
KTSARYMELVGMVLDKGTSEKOWIQEDQASYISFNAASNSRSQIKALDNAGKIMSL  
TKTAPDYLVGOQVEDISSNRIYKILELNGYDPOYASVFLGWATKKFGKRNITWLFG  
PATGKTINIAEAIHAHTVPFGCVNMTNENFPENDCVDKMVIWEEGKMTAKVESAKA  
ILGSKVRVDQCKSSAQIDPTPIVITSNTMCAVIDNSTTFHQOPLQDRMEKREL  
TRRLDHFQKVTQOEKDFEFMAKDHVEVEHEFYVKKGAKKRPAAPSDADISEPKRV  
RESVAQSTSDAEASINYADRLARGHSL"  
CDS 321. .2186  
/codon\_start=1  
/product="Rep 78 protein"  
/protein\_id="AAC03775.1"  
/db\_xref="GI:2906018"  
/translation="MPGFYEIVIKVPSDLDEHLPGISDSFVNVAEKEMELPPDSMD  
LNLIQAPLTVAEKIQDFLEMRVRVSKAPEALFVQFEKESYFHMVLVETGKXS  
MVLGRFLSQIREKLIQRIYRGIEPTLPNFAVTKTRNGAGGNKVDECYIPNYLLPK  
TOPELQAWTMEQYLACNLNTERKRLVAQHLTHVSQTOENKENONPNSDAPYTRS  
KTSARYMELVGMVLDKGTSEKOWIQEDQASYISFNAASNSRSQIKALDNAGKIMSL  
TKTAPDYLVGOQVEDISSNRIYKILELNGYDPOYASVFLGWATKKFGKRNITWLFG  
PATGKTINIAEAIHAHTVPFGCVNMTNENFPENDCVDKMVIWEEGKMTAKVESAKA  
ILGSKVRVDQCKSSAQIDPTPIVITSNTMCAVIDNSTTFHQOPLQDRMEKREL  
TRRLDHFQKVTQOEKDFEFMAKDHVEVEHEFYVKKGAKKRPAAPSDADISEPKRV  
RESVAQSTSDAEASINYADRLARGHSL"  
variation 370  
/note="compared to sequence of GenBank Accession Number  
J01901"  
/replace="g"  
precursor\_RNA 878. .4451  
CDS join(993. .1906,2228. .2252)  
/codon\_start=1  
/product="Rep 40 protein"  
/protein\_id="AAC03776.1"  
/db\_xref="GI:2906019"  
/translation="MELVGMVLDKGTSEKOWIQEDQASYISFNAASNSRSQIKALD  
NAGKIMSLTKTAPDYLVGOQVEDISSNRIYKILELNGYDPOYASVFLGWATKKFGK  
RNTIWLFGPATGKTINIAEAIHAHTVPFGCVNMTNENFPENDCVDKMVIWEEGKMTA  
KVESAKAIILGSKVRVDQCKSSAQIDPTPIVITSNTMCAVIDNSTTFHQOPLQ  
DRMEKRELTRRLDHFQKVTQOEKDFEFMAKDHVEVEHEFYVKKGAKKRPAAPSDA  
DISEPKRVRESVAQSTSDAEASINYADRLARGHSL"  
CDS 993. .2186  
/codon\_start=1  
/product="Rep 52 protein"  
/protein\_id="AAC03777.1"  
/db\_xref="GI:2906020"

/translation="MELVGMVLDKGTSEKOWIQEDQASYISFNAASNSRSQIKALD  
NAGKIMSLTKTAPDYLVGOQVEDISSNRIYKILELNGYDPOYASVFLGWATKKFGK  
RNTIWLFGPATGKTINIAEAIHAHTVPFGCVNMTNENFPENDCVDKMVIWEEGKMTA  
KVESAKAIILGSKVRVDQCKSSAQIDPTPIVITSNTMCAVIDNSTTFHQOPLQ  
DRMEKRELTRRLDHFQKVTQOEKDFEFMAKDHVEVEHEFYVKKGAKKRPAAPSDA  
DISEPKRVRESVAQSTSDAEASINYADRYONKSRHVGMNMLFPCRQERMONNSNICFTHGOK  
DLECFPVSESQPVSVKKAQYKLCYIHHIMGKVPDACTACDLVNVLDLDCIFEQ"  
precursor\_RNA 1853. .4451  
intron 1907. .2227  
CDS 2203. .4410  
/codon\_start=1  
/product="major coat protein VP1"  
/protein\_id="AAC03780.1"  
/db\_xref="GI:2906023"  
/translation="MADGYLPDWLEDTLSEGIROMWKLKPGPPPKPAERHKDSDRG  
LVLPGYKYLPGFNGLDKGEPPVNEADAALAEHDKAYDRQLDSGDNPLYLKYNHADAEFOE  
RLKEDTSFGNLRGAVFOAKKRVLEPLGLVEEPYKTAPEGKRPVEHSVPEPDSSSGTG  
KAGQOPARKRLNFGQTDGADSVDPDPOPLGPPAPSGLTNTMATSGAPMADNNEGA  
DVGNSSGNMHCSTWNGDRVITTTSTRTVALPTYNHLYKQISSQSGASNDNHYFGYS  
TPWGYEDNFRFCHFSRDMQRLINNMWGFPRKLNFKLENIQVEKTEVONDTGTTIAN  
NLTSVQVFTDSEYQLPYVLGSAHQGCLPPFPADVFMVPOYGYLTLNNGSQAVGRSSF  
XCLEYFERSQMLRTGNNTFSYTFEDVPFSSYAHQSOLDRLNPLIDQYLYLSRTNT  
PSGTTQSLQFSQAGASDIRDOSRNLPGPCYRQOQSVKTSADNNSEYSGATKY  
HLNGRDSLVPNGPAMASHKDEKEFPQSGVILFGKQSEKTNVDIEKMITDEEIR  
TNPVATPELQGSVSTNLQNGNROATADVNTQGVLPGMWQDRDVLQGPIMAKIPHT  
DGHFHPSPLMGSGFGLKHPPOILIKNTVPANPSTTFSAAKFASFITQYSTGVSVET  
EMELQKENSKRWNPEIQYTSNYSKSVNVDFTVDTNGVYSEPRPIGTRYLTRNL"  
2429  
/note="compared to sequence of GenBank Accession Number  
J01901"  
/replace="ta"  
2614. .4410  
/codon\_start=1  
/transl\_except=(pos:2614. .2616, aa:Met)  
/product="major coat protein VP2"  
/protein\_id="AAC03778.1"  
/db\_xref="GI:2906021"  
/translation="MAPGKRPRVESHSPVEPDSSSGTGKAGQOPARKRLNFGQTDADS  
VPDPOPLGPPAPSGLTNTMATSGASGAPMADNNEGADVNGSSGDNFCHFSRDMWO  
ITTSTRTVALPTYNHLYKQISSQSGASNDNHYFGYSTPWGYFDNRFCHFSRDMWO  
RLINNMWGFPRKLNFKLENIQVEKTEVONDTGTTIANLTSVQVFTDSEYQLPYVLG  
SAHQGCLPPFPADVFMVPOYGYLTLNNGSQAVGRSSFYCLEYFERSQMLRTGNNTFSY  
TFEDVPFSSYAHQSOLDRLNPLIDQYLYLSRTNTPSGTTQSLQFSQAGASDIR  
DOSRNLPGPCYRQOQSVKTSADNNSEYSGATKYHLNGRDSLVPNGPAMASHKDD  
EKEFPQSGVILFGKQSEKTNVDIEKMITDEEIRITNPVATEQYGSVSTNLQNGN  
ROATADVNTQGVLPGMWQDRDVLQGPIMAKIPHTDGHFHPSPLMGSGFGLKHPPO  
ILIKNTVPANPSTTFSAAKFASFITQYSTGVSVETEMELQKENSKRWNPEIQYTSN  
YKNSVNVDFTVDTNGVYSEPRPIGTRYLTRNL"  
2809. .4410  
/codon\_start=1  
/product="major coat protein VP3"  
/protein\_id="AAC03779.1"  
/db\_xref="GI:2906022"  
/translation="MATGSGAPMADNNEGADVNGSSGNMHCSTWNGDRVITTTSTR  
TVALPTYNHLYKQISSQSGASNDNHYFGYSTPWGYFDNRFCHFSRDMQRLINNMW  
GFRPKLNFKLENIQVEKTEVONDTGTTIANLTSVQVFTDSEYQLPYVLGSAHQGCL  
PPFPADVFMVPOYGYLTLNNGSQAVGRSSFYCLEYFERSQMLRTGNNTFSYTFEDVPF  
HSSYAHQSOLDRLNPLIDQYLYLSRTNTPSGTTQSLQFSQAGASDIRDOSRNL  
PGPCYRQOQSVKTSADNNSEYSGATKYHLNGRDSLVPNGPAMASHKDEKEFPQ  
SGVILFGKQSEKTNVDIEKMITDEEIRITNPVATEQYGSVSTNLQNGNROATAD  
VNTQGVLPGMWQDRDVLQGPIMAKIPHTDGHFHPSPLMGSGFGLKHPPOILIKNT  
VPANPSTTFSAAKFASFITQYSTGVSVETEMELQKENSKRWNPEIQYTSNYSKSVN  
VDFTVDTNGVYSEPRPIGTRYLTRNL"  
2877  
/note="compared to sequence of GenBank Accession Number  
J01901"  
/replace="c"  
3759. .3765  
/note="compared to sequence of GenBank Accession Number  
J01901"  
/replace="g"  
3859



/note="compared to sequence of GenBank Accession Number J01901"  
/replace="a"  
3898  
variation /note="compared to sequence of GenBank Accession Number J01901"  
/replace=""  
3900 .3902  
variation /note="compared to sequence of GenBank Accession Number J01901"  
/replace="gaac"  
4232 .4233  
variation /note="compared to sequence of GenBank Accession Number J01901"  
/replace="acg"  
4329 .4330  
variation /note="compared to sequence of GenBank Accession Number J01901"  
/replace="tcg"  
4336  
variation /note="compared to sequence of GenBank Accession Number J01901"  
/replace=""  
4341  
variation /note="compared to sequence of GenBank Accession Number J01901"  
/replace="c"  
4347  
variation /note="compared to sequence of GenBank Accession Number J01901"  
/replace="t"  
4535 .4679  
repeat\_region /note="inverted terminal repeat"  
/rpt\_type=inverted  
4597 .4638  
misc\_feature /note="flop oriented DNA"  
BASE COUNT 1198 a 1262 c 1255 g 964 t  
ORIGIN

Query Match 65.9%; Score 3107.8; DB 14; Length 4679;  
Best Local Similarity 80.2%; Pred. No. 0;  
Matches 3790; Conservative 0; Mismatches 882; Indels 53; Gaps 10;

QY 1 TTGCCACTCCTCTCTCTGCGCGCTCGCTCGGTGGGCGCTGCGCAAAAGTCCGC 60  
DB 1 TTGGCCACTCCTCTCTCTGCGCGCTCGCTCGTCACTGAGCGCGCGCAAAAGTCCGC 60

QY 61 AGACGGCAGAGCTCTGCTCTGCGCGCGCGCGCGCGCGCGCGCGCGAGAGGAGTG 120  
DB 61 CGACGCGCGGCTTGGCCCGCGCGCTCACTGAGCGAGCGCGCGAGAGGAGTG 120

QY 121 GGCAACTCCATCACTAGGGGTAATCGGAAGCGCTCCCAAGCTGCCGCTCAGCGCTGA 180  
DB 121 GCCAACTCCATCACTAGGGGT-----TCCTGAGAGGGGTGAGTCTGA 163

QY 181 CGTAATTACGTATAGG---GAGTGTCTGTATTAAGTGTCAAGTGTCTTTGC 237  
DB 164 CGTAATTACGTATAGGGTTAGGAGGTCTGTATTAAGGTCAAGTGTG-TTTTGC 222

QY 238 GACATTTTGGCAGACCAAGTGGCCATTAGGTAATATATGGCCGAGTGAAGCAGAGAT 297  
DB 223 GACATTTTGGCAGACCAAGTGTGTCAAGCTGGTAATTAAGCCCGAGTGAAGCAGAGGT 282

QY 298 CTCCATTTTG-AACCGGAAATTGAACGAGCAGACCATGCGGGCTTCTACGAGATCG 356  
DB 283 CTCCATTTTGAAGCGGAGGTTGAACGCGCAGCCCATGCGGGTTTACGAGATTG 342

QY 357 TGATCAAGGTGCGAGCGACTGACGAGCACTGCCGGCATTTCTGACTGTTTGTGA 416  
DB 343 TGATTAAGGTGCCAGCGACCTTGAAGCACTGCCGGCATTTCTGACAGCTTTGTGA 402

QY 417 GCTGGGTGCGGAGAGGAATGGAGCTGCCCGGATTTGACATGATCTGAATCTGA 476  
DB 417 GCTGGGTGCGGAGAGGAATGGAGCTGCCCGGATTTGACATGATCTGAATCTGA 476

DB 403 ACTGGGTGCGCGAGAGGAATGGAGTTGCCGCCAGATTCTGACATGATCTGAATCTGA 462

QY 477 TTGAGCAGCACCCCTTGACCGGTGGCGGAGAGCTGACGGCGACTTCTGTGCTCAATGGC 536  
DB 463 TTGAGCAGCACCCCTTGACCGGTGGCGGAGAGCTGACGGCGACTTCTGTGACGAATGGC 522

QY 537 GCCCGGTGAGTAAGGCCCGGAGGCCCTCTTCTTTGTTCAAGTTCCGAGAGGCGAGTCT 596  
DB 523 GCCGTGTGAGTAAGGCCCGGAGGCCCTTCTTCTTGTGCAATTTGAGAAGGAGAGAGCT 582

QY 597 ACTTCCACTCCATATTTCTGTGTGAGACCAACGCGGGGTCAATCCATGCTGTGGCGCT 656  
DB 583 ACTTCCACTGACAGCTGCTGTGGAACCAACGCGGGGTGAATCCATGCTTTGGAGCGTT 642

QY 657 TCCTGAGTCAGATTAGGACCAAGCTGTGACAGCCATCTACCGCGGATCGAGCCGACCC 716  
DB 643 TCCTGAGTCAGATTGCGGAAAACTGATTAGAGAAATTACCGCGGATCGAGCCGACTT 702

QY 717 TGCCCACTGTTTGGGTGACCAAGCGCTAATGCGCGGAGGGGGAACAAGGTGG 776  
DB 703 TGCCCACTGTTTGGGTGACCAAGCGCTAATGCGCGGAGGGGGAACAAGGTGG 762

QY 777 TGACGAGTGCTACATCCCAACTACCTCTGCCCAAGACTGACCCGAGCTGACGTGG 836  
DB 763 TGATGAGTGCTACATCCCAACTACTGTCTCCCAAAACCGAGCTGAGCTCCAGTGG 822

QY 837 CGTGACTAATGAGAGGATATTAAGCGCTGTTGAACCTGGCCGAGCGCAACGGC 896  
DB 823 CGTGACTAATGAGAGGATATTAAGCGCTGTTGAATCTCACGAGCGTAACGGT 882

QY 897 TCGTGGCGCAGCACCCTGACCCCACTGACCCAGACCCAGAGAGCAACAAGGAATCTGA 956  
DB 883 TGTGGCGCAGCATCTGACCGCAGCTGTCCAGACGAGAGAGCAACAAGGAATCTGA 942

QY 957 ACCCAATCTGACGCGCTGTGATCCGGTCAAAAACTCCGCGCTACATGAGCTGG 1016  
DB 943 ATCCCAATCTGATGCGCGGTGATGATCAAAAACTTCAAGCAAGTACATGAGCTGG 1002

QY 1017 TCGGTGCTGTGACCGGGGCACTCACTCCGAGAGCAGTGAATCCAGAGAGCAGG 1076  
DB 1003 TCGGTGCTGTGACCAAGGGGATTACTCCGAGAGCAGTGAATCCAGAGAGCAGG 1062

QY 1077 CCTGTACATCTCCTTCAACGCGCTTCCAACTCGCGGTCCAGATCAAGCGCTGTGG 1136  
DB 1063 CCTGTACATCTCCTTCAATGCGGCTCCAACTCGCGGTCCAAATCAAGGCTGCTTGG 1122

QY 1137 ACAATGCGGCAAGATCATGCGCTGACCAATCCGCGCGCACTACCTGTAGGCCCG 1196  
DB 1123 ACAATGCGGAAAGATTAAGCTGACTAAACCGCGCGCACTACCTGTGGGCCAGC 1182

QY 1197 CTCGCGCGCGGACATTAACCAACCAACCGCATCTACCGCATCTGAGCTGAACGGCTACG 1256  
DB 1183 AGCCCGTGAAGACATTTCCAGCAATCGGATTTATAAATTTTGAACCTAAACGGGTACG 1242

QY 1257 AACCTGCTACGCGGCTCGTCTTCTCGGCTGGGCCCAAGAAAGGTTGGAAGCGCA 1316  
DB 1243 ATCCCAATATGCGGCTTCCGCTTCTGGAATGGCCACGAAAGTTGCGCAAGAGGA 1302

QY 1317 ACACATCTGCTGTTTGGCGCGGCAACGCGGCAAGCAACATCGCGGAAGCCATCG 1376  
DB 1303 ACACATCTGCTGTTTGGCGCTGCACTACCGGAAGACCAACATCGCGGAGCCATAG 1362

QY 1377 CCCAGCGGTGCCCTTCTACGCGTGGTCACTGACCAATGAACCTTCCCTCAATG 1436  
DB 1363 CCCAGCTGTGCCCTTCTACGCGTGGTAACTGACCAATGAACCTTCCCTCAACG 1422

QY 1437 ATTGCGTGAACAAGTGTGATCTGTGGAGGAGGGAAGTGAACGCGCAAGTCTGG 1496  
DB 1423 ACTGTGTGAACAAGTGTGATCTGTGGAGGAGGGAAGTGAACCGCAAGTCTGG 1482

QY 1497 AGTCCGCAAGGCCATTTCTGGCGGCAAGAGTGGCGGTGACCAAAAGTGAAGTCTGT 1556  
DB 1483 AGTCCGCAAGGCCATTTCTGGAGAGCAAGGTGGCGGTGACCAAAATGCAAGTCT 1542

QY	1557	CCGCCAGATCGACCCCAACCCCGTGATGTCACCTTCAACACCAACATGTGCGCGGTGA	1616
Db	1543	CGGCCAGATAGACCCGACTCCCGTGATGTCACCTTCAACACCAACATGTGCGCGGTGA	1602
QY	1617	TTGACGGGAACAGCACCACTTTCAGACCAACAGCAGCCGTTGCAGGACCCGATGTTCAAT	1676
Db	1603	TTGACGGGAACCTCAACGACTTTCGAACACCAAGCAGCCGTTGCAGGACCCGATGTTCAAT	1662
QY	1677	TTGAACCTACCCCGCTGCGAGCATGACTTTGGCAAGGTGACAAAGCAGGAAGTCAAAG	1736
Db	1663	TTGAACCTACCCCGCTGCGAGCATGACTTTGGGAAGGTCAACCAAGCAGGAAGTCAAAG	1722
QY	1737	AGTTCTTCCGCTGGGCGCAGATCACTGACCGAGTGGCCCATGAGTTCTACGTCAGAA	1796
Db	1723	ACTTTTCCGCTGGGCGCAAGGATCACTGAGTTGAGGTGAGCATGAATTTCTACGTCAGAA	1782
QY	1797	AGGGTGGAGCCAAACAAAGACCCGCCCGATGACCGCGGATAAAGCGAGCCCAAGCGGG	1856
Db	1783	AGGGTGGAGCCAAAGAAAGACCCGCCCGAGTGACGAGATATTAAGTGAGCCCAACGGG	1842
QY	1857	CCTGCCCTCAGTCCGCGATCCATCGACGCTCAGACCGCGAAAGAGCTCCGGTGACTTTG	1916
Db	1843	TGCGCGAGTCAGTTGCGCAGCATCGACGTCAGACGGCGA--AGCTTGATCAACTACG	1899
QY	1917	CCGACAGGTACCAAAACAAATGTTCTCGTCAACGGCGCATGCTTCAGATGCTGTTCCCT	1976
Db	1900	CAGACAGGTACCAAAACAAATGTTCTCGTCAACGTCGCGCATGAATCTGATGCTGTTCCCT	1959
QY	1977	GCAAGACATGCGAGAGATGAATCAGAATTTCAACATTTGCTTCAACGACCGGAGAGAG	2036
Db	1960	GCAAGACATGCGAGAGATGAATCAGAATTTCAAAATATCTGCTTCACTCAGCAGAGAAAG	2019
QY	2037	ACTGTTCAGAGTGTCTCCCGCGCTGTCAGAATCTCAACCGGT--CGTCAGAAAGAGA	2093
Db	2020	ACTGTTAGAGTGTCTTCC--CGTGTCAGAATCTCAACCCGTTTCTGTGTCAAAAAG	2076
QY	2094	CGTATCGGAAACTCTGTGCCATTCATCATCTGTCGGCGGGCTCCCGAGATTGCTTGCT	2153
Db	2077	CGTATCAGAAACTGTGTCTATCATTCATCATATCATGGGAAAGGTGCCAGA--CGCTTGCA	2133
QY	2154	CGGCTCGCATCTGGTCAACGTGGAACCTGATGACTGTGTTCTGAGCAATAATGACTT	2213
Db	2134	CTGCCTCGATCTGGTCAATGTGCAATTTGGATGACTGCACTTTGAACAATAATGATTT	2193
QY	2214	AAACCAAGTATGGCTGCCATGGTTATCTTCCAGATTGGCTCGAGACAACCTCTCTGAG	2273
Db	2194	AAATCAGGTATGGCTGCCATGGTTATCTTCCAGATTGGCTCGAGACAACCTCTCTGAA	2253
QY	2274	GGCATTCGCGAGTGTGGCACTTGAACCTGAGACCCCGAAAGCCCAAGCCAACGACAA	2333
Db	2254	GGAATTAAGACAGTGTGGAGCTCAAACTGAGCCCAACCAACCAAGCCGCAAGCGG	2313
QY	2334	AAGCAGGACGACGGCGGGGCTGTGCTCTTGGCTACAAGTACCTCGGACCTTCAAC	2393
Db	2314	CATAAGGACGACGACGGGCTTGTGCTTCTTGGGTACAAGTACCTCGGACCTTCAAC	2373
QY	2394	GGACTCGACAAAGGGGAGCCCGTCAACGAGGCAAGCCGCGCCCTCGAGCAGACAA	2453
Db	2374	GGACTCGACAAAGGGAGAGCCCGTCAACGAGGCAAGCCGCGCCCTCGAGCAGACAA	2433
QY	2454	GCCTACGACCAAGGGGAGCCCGTCAACGCGCGGAGCGGAGCGGCCCTCGAGCAGACA	2513
Db	2434	GCCTACGACCGGAGCTGACACGAGACAAACCGTCAAGTACAACCAACGACCGAC	2493
QY	2514	GCCGAGTTTCAGAGCGTCTGCAAGAAGATACGCTTTTGGGGCAACCTCGGGGAGCA	2573
Db	2494	GCCGAGTTTCAGAGCGCCTTAAGAAGATACGCTTTTGGGGGCAACCTCGGAGAGCA	2553
QY	2574	GTCTTCCAGGCCAAGAGCGGGTTCTCGAACCTCTCGGTCTGTTGAGGAAGGCGCTAAG	2633
Db	2554	GTCTTCCAGGCCAAGAGAGGGTTCTTGAACCTCTGGGCTGTTGAGGAACCTGTTAAG	2613

QY	2634	ACGGCTCCTGGAAAGAAACGTCCGGTAGAGCAGTCCGCACAAGAGCCAGACTCTCTCG	2693
Db	2614	ACGGCTCCGGGAAAAAAGAGGCCCGGTAGAGCACTCTCTGTGAGCCAGACTCTCTCG	2673
QY	2694	GGCATCGGCAAGACAGGCCAGCAGCCCCGTAAAAAGACTCAATTTTGGTCAGACTGGC	2753
Db	2674	GGAAACCGAAAGCCGGCCAGCAGCCTGCAAGAAAAAGATTGAATTTTGGTCAGACTGGA	2733
QY	2754	GACTCAGAGTCAGTCCCCGATCCACAACCTCTCGAGAAACCTCCAGCAACCCCGCTGCT	2813
Db	2734	GACGCAGACTCAGTACTGACCCCGAGCCTCTCGGACAGCCACAGCAGCCCCCTCTGGT	2793
QY	2814	GTGGACCTACTACATGGCTTCAGCGGTGGCGCAACATGGCAGACAATAAGAAAGCC	2873
Db	2794	CTGGGACTAATACGATGGCTACAGGCAGTGGCGCAACCAATGGCAGACATAACGAGGC	2853
QY	2874	GCCGACGAGTGGGTAAATGCTCAGGAAATTGGCAATTGCGATTCCACATGGCTGGCGAC	2933
Db	2854	GCCGACGAGTGGGTAAATTCCTCGGAAATTGGCAATTGCGATTCCACATGGATGGCGAC	2913
QY	2934	AGAGTCATCACCCACAGCACACCCGACCTGGGCTTGGCCACCTACAATAACCACTCTAC	2993
Db	2914	AGAGTCATCACCCACAGCACACCCGAACTGGGCCCTGCCACCTACAACAACCACTCTAC	2973
QY	2994	AAGCAATCTCCAGTCTTCAACGGGGCCAGCAAGCAACCACTACTTGGCTACAGC	3053
Db	2974	AAACAAATTTCCAGCCAATCA -- GGAGCCTCGAAGCAATCACTACTTTGGCTACAGC	3030
QY	3054	ACCCCTGGGGGTATTTGATTTCAACAGATTTCACTGCCACTTTTCAACCACTGACTGG	3113
Db	3031	ACCCCTGGGGGTATTTGACTTCAACAGATTTCACTGCCACTTTTCAACCACTGACTGG	3090
QY	3114	CAGCGACTCATCAACAACATTTGGGATTTCCGGCCCAAGAGACTCAACTTCAACTCTTC	3173
Db	3091	CAAGACTCATCAACAACAACCTGGGATTTCCGACCCAAAGAGACTCAACTTCAAGCTCTT	3150
QY	3174	AACATCCAAGTCAAGAGAGTCAAGCAAGATGATGAGCTCACAAACCATCGCTAATACTT	3233
Db	3151	AACATCCAAGTCAAGAGAGTCAAGCAAGATGATGATGATGATGATGATGATGATGATGAT	3210
QY	3234	ACCAGCAGGTTCAAGTCTTCTCGACTCGAGTACCAAGCTTCCGTACGTCTCGGCTCT	3293
Db	3211	ACCAGCAGGTTCAAGTCTTCTCGACTCGAGTACCAAGCTTCCGTACGTCTCGGCTCTG	3270
QY	3294	GCGACACAGGCTGCTCCCTCCGTTCCCGCGCAGTGTTCATGATTCGCAATACGCGC	3353
Db	3271	GCGACACAGGATGCTCCCGCGCTTCCAGCAGACGTCTTCATGATGATGATGATGATGATGAT	3330
QY	3354	TACCTGACGCTCAACAATGGCAGCCAGCCGTGGGACGTTTCATCTTTTACTGCTGGAA	3413
Db	3331	TACCTGACGCTCAACAACGGGAGTCAAGGACGATGAGACGCTCTTCATTTTACTGCTGGAG	3390
QY	3414	TATTTCCCTTCTGAGATGCTGAGAACGGGCAACAATTACCTTCAGCTACACCTTTGAG	3473
Db	3391	TACTTCTCTGAGATGCTGCGTACCGGAAACAATTACCTTCAGCTACACCTTTGAG	3450
QY	3474	GAAGTGCCTTTCCAGCAGCAGCTACGCGCACAGCCAGAGCCTGGACCGGCTGATGATCCT	3533
Db	3451	GACGTTCTTCCAGCAGCAGCTACGCTACAGCCAGAGTCTGACCGCTCATGATCCT	3510
QY	3534	CTCATGCACCAATACCTGTATTACCTGAAACAAGACTCAAAATCAGTCCGGAAGTGCCTAA	3593
Db	3511	CTCATGCACCAATACCTGTATTACTTGAGCAGAACAAACACTCCAAGTGAACCAACACAG	3570
QY	3594	AACAAGACTTGTGTTTAAAGCCGTGGGTCTCAGCTGGCATGTCTGTTCAAGCCAAAAC	3653
Db	3571	CAGTCAAGGCTTCAGTTTCTCAGGCCGAGCGAGTGAATTCGGGACCACTGATGAAC	3630
QY	3654	TGGCTACCTGGAACCTGTATTCGGCAGCAGCGCGTTTCTAAAAACAACAAACAGACAAC	3713
Db	3631	TGGCTTCTGGAACCTGTATTCGGCAGCAGCGAGTATCAAAAGACATCTGCGGATTAACAAC	3690
QY	3714	AACAGCAATTTTACTGGAAGTGTGCTTCAAAATATAAATCACTCAATGGCGGTGAATCCATC	3773

|||||  
Db 3691 AACAGTGAATACTGCTGGAGCTGGAGCTAACCAAGTACCCTCAATGGCGAGAGACTCTCTG 3750  
QY 3774 ATCAACCCCTGGCACTGCTATGGCTTACACAAAGACGACGAAGACAGTCTTCCCATG 3833  
Db 3751 GTGAATCCGGGCCCGCCATGGCAAGCCACAAAGACGATGAAGAAAGTTTCTCTCAG 3810  
QY 3834 AGCGGTGTCATGATTTTGGAAAAGAGAGCGCGGAGCTTCAAAACACTGATTTGACAAT 3893  
Db 3811 AGCGGGTTCTCATCTTTGGGAAGCAAGGCTCAGAAAAACAATGTGACATTTGAAAAG 3870  
QY 3894 GTCATGATTACAGACGAAGAAATTAAAGCCCTAACCTGTGGCCACCGAAAGATT 3953  
Db 3871 GTCATGATTACAGACGAAGAAATTAAAGCCCTAACCTGTGGCCACCGAAAGATT 3930  
QY 3954 GGGACCGTGGCAGTCAATTTCCAGAGCAGACAGACAGACCTCCGACCGAGATGTGCAT 4013  
Db 3931 GGTTCGTATCTACCAACCTCCAGAGAGGCAACAGACAGCTACCGCAGATGTCAAC 3990  
QY 4014 GCTATGGAGCATTACCTGGCATGTGTGGCAGATAGACGTGTACTTGCAGGGTCCC 4073  
Db 3991 ACACAAAGCGTTCTTCCAGGCATGTGTGGCAGACAGAGATGTACTTTCAGGGGCC 4050  
QY 4074 ATTTGGGCCAAATTCCTCACAAGATGACACTTTCAACCGCTCTCTTATGGCGGC 4133  
Db 4051 ATCTGGCAAGATTTCCACACGAGCAGACATTTCAACCCCTCTCCCTCATGGGTGA 4110  
QY 4134 TTTGACTCAAGAACCCGCTCTCAGATCTCATCAAAACACGCGCTGTCTGCGAAT 4193  
Db 4111 TTGGACTTAAACACCTCTCTCAGATCTCATCAAGAACACCCGCTACCTGCGAAT 4170  
QY 4194 CCTCCGCGAGTTTTCAGCTACAAAGTTTGTCTCATTCATCACCCCAATATCCACAGA 4253  
Db 4171 CCTTCGACCACTTCACTGCGGCAAGTTGTCTCTTCATCACACAGTATCCACGGA 4230  
QY 4254 CAAGTAGTGAATTTGAATGGAGCTGCAGAAAGAAACAGCAAGCGCTGGAATCCC 4313  
Db 4231 CAGGTACGCTGAGATCGAGTGGAGCTGCAGAAAGAAACAGCAACGCTGAATCCC 4290  
QY 4314 GAAGTCACTACATCCCAATTATGCAAAATCTGCCAACGTTGATTTTACTGTGACAAC 4373  
Db 4291 GAAATTCAGTACACTTCCAACTACAACAAGTCTGTTAATGTGACTTACTGTGACACT 4350  
QY 4374 AATGACTTATATGAGCCTCGCCCATTTGGCACCCTTACCTTACCCGCTCCCTGTA 4433  
Db 4351 AATGGCGTGTATCAGAGCCTCGCCCATTTGGCACCAGATACCTGACTGTAATCTGTAA 4410  
QY 4434 TTACGTGTTAATCAATAAACCGGTTGATTCGTTCACTTGAACCTTGGCTCTCTGCTT 4493  
Db 4411 TTGCTGTTAATCAATAAACCGTTAATCGTTCACTTGAACCTTGGCTCTCTGCTATT 4470  
QY 4494 CTTATCTTATCGGTTAACCATGTTATAGCTTACATTAATGCTTGGTGGCTTCCG 4553  
Db 4471 TCTTCTTATCTAGTTCCATGGCTACGTAGATAGTACATGGCGGTTAATCATTAAC 4530  
QY 4554 ATAAAGACTTACGTATCGGGTTACCCCTAGTATGATGAGTTGCCACTCCCTCTCTCG 4613  
Db 4531 TACAAGGA-----ACCCCTAGTATGAGTTGGCCACTCCCTCTCTCTCG 4574  
QY 4614 CGCTCGCTCGCTCGGTGGGGCTCGCGACCAAGGTCCGCAAGCGGAGAGCTCTGCTCT 4673  
Db 4575 CGCTCGCTCGCTCACTGAGGGCCGGGCGAACCAAGGTGCGCCGACCGCGGCTTTGCCG 4634  
QY 4674 GCCGGCCCCCAGCGAGCGAGCGCGCAGAGAGGAGTG3CAA 4718  
Db 4635 GCGCGCCTCAGTGAAGCGAGCGCGCAGAGAGGAGTGCCAA 4679

RESULT 7  
LOCUS AX205072 8698 bp DNA linear PAT 30-AUG-2001  
DEFINITION Sequence 2 from Patent WO0155361.  
ACCESSION AX205072

VERSION AX205072.1 GI:15394313  
KEYWORDS .  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 8698)  
AUTHORS Hardy, S.F.  
TITLE Recombinant aav packaging systems  
JOURNAL Patent: WO 0155361-A 2 02-AUG-2001;  
Chiron Corporation (US)  
FEATURES  
Source location/Qualifiers  
1..8698  
/organism="unidentified"  
/db\_xref="taxon:32644"  
/note="recombinant DNA"  
BASE COUNT 2117 a 2368 c 2306 g 1907 t  
ORIGIN  
Query Match. 65.9%; Score 3107.8; DB 6; Length 8698;  
Best Local Similarity 80.2%; Pred. No. 0;  
Matches 3790; Conservative 0; Mismatches 882; Indels 53; Gaps 10;  
QY 1 TTGCCCACTCCCTCTGCGCGCTCGCTCGCTCGGTGGGCGCTGCGACCAAGTCCGC 60  
Db 1 TTGGCCACTCCCTCTGCGCGCTCGCTCGCTCACTGAGCGCGCGCAGCAAGGTCCGC 60  
QY 61 AGACGGCAGAGCTCTGCTCTGCGGCGCCACCGAGCGAGCGAGCGCGCAGAGAGAGT 120  
Db 61 CGACGCGCGGCTTGGCGCGCGCTCAGTACGAGCGAGCGCGCGCAGAGAGAGT 120  
QY 121 GGCACTTCATCCTAGGGGTATGCGGAAGCGCTCCACGCTCCGCTCAGCGCTGA 180  
Db 121 GCCAATTCATCCTAGGGGTATGCGGAAGCGCTCCACGCTCCGCTCAGCGCTGA 163  
QY 181 CGTAATTTAGCTATAGG--GAGTGTCTGTATTTAGCTGTCAAGTGTCTTTG 237  
Db 164 CGTAATTTAGCTATAGGTTAGGAGGTCTGTATTAGAGTCAAGTGTG- TTTTGC 222  
QY 238 GACATTTTGCAGACACCAAGTGGCATTTAGGGTATATATGCGGAGTGAGCGAGCAGAT 297  
Db 223 GACATTTTGCAGACCAATGTGTACAGCTGCTGGTATTAAAGCCGAGTGAGCAGAGGT 282  
QY 298 CTCCATTTTG-ACCGGAATTTGAACGAGCAGCAGCAGCCATGCGGGCTTCTACGAGATCG 356  
Db 283 CTCCATTTTGAAGCGGAGGTTGAAGCGCAGCGCCATGCGGGCTTTACGAGATTTG 342  
QY 357 TGATCAAGGTGCGGAGCGACCTGACGACACCTGCGCGCATTTCTGACTCGTTGTGA 416  
Db 343 TGATTAAGTCCCGACGACCTTGAAGCATCTGCCGCAATTTCTGACAGCTTTGTGA 402  
QY 417 GCTGGTGGCCGAGAAAGATGGAGCTGCCCCCGGATTTTGACATGGATCTGAATCTGA 476  
Db 403 ACTGGGTGGCCGAGAAAGATGGAGTTGCCCGCAGATTTTGACATGGATCTGAATCTGA 462  
QY 477 TTGAGCAGGCAACCCCTGACCGGTGGCGGAGAAAGCTGACGCGGCACTTCTGTGCCAATGCC 536  
Db 463 TTGAGCAGGCAACCCCTGACCGGTGGCGGAGAAAGCTGACGCGGCACTTCTGTGCCAATGGC 522  
QY 537 GCCCGGTGAGTAAGCCCCCGAGCCCTCTTTGTTGACGTTGAGAAGGGCGAGTCT 596  
Db 523 GCCGTGTGAGTAAGCCCCCGAGCCCTTTTCTTTGTGCAATTTGAGAAGGAGAGAGCT 582  
QY 597 ACTTCACTTCATATTTCTGTGTGAGACCAAGGGGTCAAAATCCATGTGCTGGCGCT 656  
Db 583 ACTTCACTTCAGAGTCTGTGTGAGAAACCAAGGGGTCAAAATCCATGTGTTGGAGCTT 642  
QY 657 TCCTGAGTCAGATTAGGACAAAGCTGTGACAGACCATCTACCGCGGATCGAGCCGACCC 716  
Db 643 TCCTGAGTCAGATTGCGGAAAACTGATTCAGAGATTTACCGCGGATCGAGCCGACTT 702  
QY 717 TGCCCACTGTGTTGCGGTGACCAAGCGCTAATGCGCGCGGAGAGGAGCAAGGTGG 776  
Db 703 TGCCCACTGTGTTGCGGTGACCAAGCAAGAAATGGCGCGGAGAGGAGCAAGGTGG 762



QY	777	TGACGAGTGCTACATCCCCAACTACTCTCTGCCCCAAGACTCAGCCCGAGCTGCAGTGG	836
Db	763	TGATGAGTGCTACATCCCAAATTACTTGCTCCCCAAAACCAGCTGAGCTCCAGTGG	822
QY	837	CGTGACTTAACATGAGAGATATATAAGCGCTGTTTGAACCTTGCCGAGCGCAACGCG	896
Db	823	CGTGACTAATATAGAACAGATATTAAAGCGCTGTGTTGAATCTCACGGAGCGTAACGCT	882
QY	897	TCGTGGCGCAGCACCTGACCCACGTCAAGCCAGACCCAGAGCAGAACAAAGGAATCTGA	956
Db	883	TGCTGGCGCAGCATCTGACGCGACGTGTCCGACAGCCAGAGCAGAACAAAGAAATCAGA	942
QY	957	ACCCCAATTCTGACCGCGCTGTCACTCCGGTCAAAAACTCCGCGCTACATGAGCTGG	1016
Db	943	ATCCCAATTCTGATGCGCGGTGATCAGATCAAAAACTTACGCCAGGTACATGAGCTGG	1002
QY	1017	TCGGGTGCTGTGAGACCGGGCATCACTCCGAGAAAGCATGATCCAGAGAGACCAAG	1076
Db	1003	TCGGGTGGCTCGTGACAAAGGGATTAACCTCGAGAAAGCATGATCCAGAGAGACCAAG	1062
QY	1077	CCTCGTACATCTCCTTCAACGCGCTTCCAACTCGCGGTCCAGATCAAGGCGCTTGG	1136
Db	1063	CCTCATACATCTCCTTCAATGGCGCTCCAACTCGCGGTCCCAATCAAGGCTGCTTGG	1122
QY	1137	ACAATGCGGCAAGATCATGGCGCTGACCAAAATCCGCGCCGACTACCTGTAGGCCCG	1196
Db	1123	ACAATGCGGAAAGATTATGAGCTGACTAAAAACGCGCCGACTACCTGTGGGCCAGC	1182
QY	1197	CTCCGCGCGGACATTTAAAAACAACCGCATCTACCGCATCTGAGCTGAACGGCTACG	1256
Db	1183	AGCCGTGGAGACATTTCCAGCAATCGGATTATAAAAATTTGAACTAAACGGGTACG	1242
QY	1257	AACTGCTACGCGCGCTCCGTCTTCTCGGCTGGGCCAGAAAAGTTGCGGAAGCGCA	1316
Db	1243	ATCCCCAATATAGCGGCTTCCGTCTTCTGCGATGGGCCCAAGAAAAGTTGCGCAAGAGA	1302
QY	1317	ACAACATCTGGCTGTTGGGCGGSCCAACACGGGCAAGACCAATCGCGGAAGCCATCG	1376
Db	1303	ACAACATCTGGCTGTTGGGCTTGCAACTACCGGGAAGACCAATCGCGGAGGCCATAG	1362
QY	1377	CCCAACGCGTGCCCTTCTACGGCTGCGTCACTGACCAATGAGAACTTTCCTTCAATG	1436
Db	1363	CCCAACATGTCCTTCTACGGGTCGTAACCTGACCAATGAGAACTTTCCTTCAACG	1422
QY	1437	ATTGCTGCAACAAGATGGTGAATCTGGTGGAGAGAGGCAAGATGACGGCCAAAGTCTGG	1496
Db	1423	ACTGTGTCGACAAGATGGTGAATCTGGTGGAGAGAGGGAAGATGACCGCCAAAGTCTGG	1482
QY	1497	AGTCCGCCAAAGCCATTCTCGGCGGCAAGCAAGGTGCGCGTGAACCAAAAGTGCAAGTCT	1556
Db	1483	AGTCCGCCAAAGCCATTCTCGAGGAAGCAAGGTGCGCGTGAACCAAAATGCAAGTCT	1542
QY	1557	CCGCCAGATGACCCCAACCCCGTGAATGTCACCTTCAACCAACATGTGCGCCGTA	1616
Db	1543	CGGCCAGATAGACCCGACTCCGTGATGTCACCTTCAACCAACATGTGCGCCGTA	1602
QY	1617	TTGACGGGAACAGCACCACTTGAGCACCAAGCAGCGCTTGACGACCAGATGTTCAAT	1676
Db	1603	TTGACGGGAACCTCAACGACTTGACAACACAGCAGCGCTTGACAAGACCGATGTTCAAT	1662
QY	1677	TTGAATCTACCCCGCTTGAGCATGACTTTGGCAAGTTCACAAAGCAGGAAGTCAAG	1736
Db	1663	TTGAATCTACCCCGCTTGAGTCACTACTTTGGGAAGTTCACCAAGCAGGAAGTCAAG	1722
QY	1737	AGTTCTTCCGCTGGGCGCAGATCACTGACCGAGGTGGCGCATGAGTTCTACGTCAAG	1796
Db	1723	ACTTTTCCGGTGGCAAAAGATCACTGTGTTGAGGTGAGCATGAATTTCTACGTCAAA	1782
QY	1797	AGGTTGAGCCCAACAAAGACCCGCCCGCATGACGCGGATTAAGCGAGCCCAAGCGG	1856
Db	1783	AGGTTGAGCCCAAGAAAGACCCGCCCGCATGACGCAAGTATTAAGTGAGCCCAAGCGG	1842

QY	1857	CCTGCCCTCAGTCCGGATCCATCGACGTGAGACCGGAAGAGCTCCGGTGACTTTG	1916
Db	1843	TGCGGAGTCA GTTGCAGCCATCGACGTGAGCGGGA -- AGCTTGATCAACTACG	1899
QY	1917	CCGACAGGTACCAAAACAATGTCTTCGTACAGCGGGCATGCTTCAGATGCTGTTCCCT	1976
Db	1900	CAGACAGTACCAAAACAATGTCTTCGTACAGTGGGCATGAATCTGATGCTGTTCCCT	1959
QY	1977	GCAAGACATGCGAGAGATGAATCAGAATTTCAACAATTGCTTCACGACGGAGCAGAG	2036
Db	1960	GCAGACATGCGAGAGATGAATCAGAATTTCAAAATCTGCTTCACTCACGACAGAAAG	2019
QY	2037	ACTGTTCAAGTGTCTCCCGCGGTGTCAGAATCTCAACCGT -- CGTCAGAAAGAGA	2093
Db	2020	ACTGTTAGAGTGTCTTCC -- CGTGTCAGAATCTCAACCCGTTTCTGTCTCAAAAGG	2076
QY	2094	CGTATCGGAACTCTGTGCCATTTCATCATCTGCTGGGCGGGCTCCCGAGATTGCTTGT	2153
Db	2077	CGTATCAGAACTGTGTCAATTCATCATATCATGGAAGGTGCCAGA -- CGCTTGCA	2133
QY	2154	CGGCTCGATCTGTCAACGTGACCTGATGACTGTGTTCTGAGCAATAAATGACTT	2213
Db	2134	CTGCTCGATCTGTCAATGTGATTGGATGACTCATCTTGAAACAATAATGATTT	2193
QY	2214	AAACAGGTATGGCTGCGGATGTTATCTTCCAGATTGGCTCGAGACAACCTCTGAG	2273
Db	2194	AAATCAGGTATGGCTGCCGATGTTATCTTCCAGATTGGCTCGAGACACTCTCTGAA	2253
QY	2274	GGCATTCCGAGTGTGTGGACTTGAACCTTGAGCCGCCGAAGCCCAAGCCACACGAA	2333
Db	2254	GGAATAGACAGTGTGTGAAGCTCAACCTGGCCCAACCAACCAAGCCCGCAGAGCGG	2313
QY	2334	AAGCAGGACGACGGCCGGGCTGTGTCTTCTGTGCTCAAGTACCTCGGACCTTCAAC	2393
Db	2314	CATAAGGACGACGAGGGGCTTGTGCTTCTGTGGTACAAGTACCTCGGACCTTCAAC	2373
QY	2394	GGACTCGACAAGGGGAGCCCTCAACGCGCGGACGACGCGCCCTCGAGCAGCAAG	2453
Db	2374	GGACTCGACAAGGGAGAGCCGTCAACGAGCAGACGCGCCCTCGAGCAGCAAA	2433
QY	2454	GCCTACGACCAGCAGCTCAAAAGCGGTGACCAATCCGTACTGCGTATAACACGCGCAC	2513
Db	2434	GCCTACGACCAGCAGCTCGACAGCGAGACAACCCGTACTCAAGTACAACACGCGCAC	2493
QY	2514	GCCGAGTTTCAGAGCGTCTGCAAGAAGATACGTTTTTGGGGGCAACCTCGGGCGACA	2573
Db	2494	GCCGAGTTTCAGAGCGCTTAAAGAAGATACGTTTTTGGGGGCAACCTCGGACGACA	2553
QY	2574	GTCCTTCCAGGCCAAGAACGGGTTCTCGAACCTCTGSGTCTGTTGAGGAAGCGCTAAG	2633
Db	2554	GTCCTTCCAGGCCAAGAAAGGGGTTCTGAACCTCTGGGCGCTGTTGAGGAACCTGTTAAG	2613
QY	2634	ACGGCTCCTGGAAGAAAGTCCGGTAGAGCAGTCGCCACAAGAGCCAGACTCCTCTCG	2693
Db	2614	ACGGCTCCGGGAAGAAAGAGCGCGTAGAGCACTCTCTGTGAGCCAGACTCCTCTCG	2673
QY	2694	GGCATCGGCAAGACAGGCCAGACCCGCTAAAGAGACTCAATTTTGTGTCAGACTGCG	2753
Db	2674	GGAACCGGAAGCGGGGCCAGACGCTCGAAGAAAGATTGAATTTTGTGTCAGACTGGA	2733
QY	2754	GACTCAGAGTCACTCCCGATCCACAACCTCTCGGAGAACCTCCAGCAACCCCGCTGCT	2813
Db	2734	GACGCAAGTCACTCACTGACCCCGAGCTCTCGGACAGCCACAGACGCCCTCTGCT	2793
QY	2814	GTGGACCTACTACAAATGGCTTCAGGCGGTGGCGCACCAATGGCAGACAATAAGCAGC	2873
Db	2794	CTGGAACTAATACGATGGCTACAGGCAGTGGCGCACCAATGGCAGACAATAAGCAGC	2853
QY	2874	GCCGACGAGTGGGTAAATGCTCAGGAAATTTGGCAATTCCAGATGCGTGGGCGAC	2933
Db	2854	GCCGACGAGTGGGTAAATCTCTCGGAAATTTGGCAATTCCAGATGAGTGGGCGAC	2913
QY	2934	AGAGTCATCACCAAGCACCCGACCTGGGCTTGCCCACTCAATAACCACTTAC	2993

Db	2914	AGAGTCATCACCA	CAGCACCAGAACTGGGCCCTGCCCACTCAACAACACCACTTAC	2973
QY	2994	AAGCAAATCTCC	AGTGTCTCAACGGGGGCCAGCAACGAACTACTTCGGCTACAGC	3053
Db	2974	AAACAATTTCC	AGCCAATCA---GGAGCCTCGAACGACATCACTACTTTGGCTACAGC	3030
QY	3054	ACCCCTGGGGG	TATTTTGATTTTCAACAGATTCCACTGCCACTTTTCAACCAGTGACTGG	3113
Db	3031	ACCCCTTGGGG	GTATTTTGACTTTCAACAGATTCCACTGCCACTTTTCAACCAGTGACTGG	3090
QY	3114	CAGCGACTCAT	CAACAACAATTGGGGATTCCGGGCCCAAGAGACTCAACTTCAAACTCTTC	3173
Db	3091	CAAGA	CTCATCAACAACACTGGGGATTCCGACCCCAAGAGACTCAACTTCAAGCTCTTT	3150
QY	3174	AACATCCAAGT	CAAGAGTCAAGCAATGATGGCGTCAACCATCGTAAATAACTT	3233
Db	3151	AACATTCAAGT	CAAGAGTCAAGCAATGACGTACGACGACGATGGCCAATTAACCTT	3210
QY	3234	ACCAGCAGGTT	CAAGTCTTCTCGGACTCGAGTACCAAGCTTCCGTACGTCCTGGCTCT	3293
Db	3211	ACCAGCAGGTT	CAAGTCTTCTCGGACTCGGAGTACCAAGCTTCCGTACGTCCTGGCTCT	3270
QY	3294	GCGCACGAGG	CTGCCCTCCCTCCGTTCCCGCGGAGCGTTCATGATTCGCCAATACGC	3353
Db	3271	GCGCATCAAG	ATGCTCTCCCGCGTTCCCAAGACAGCGTCTTATGTGTGCCACAGTATGA	3330
QY	3354	TACCTGACGCT	CAACAATGGCAAGCCCAAGCCGTGGACGTTCACTCTTTTACTGCTCGAA	3413
Db	3331	TACCTCA	CCCTGAACAAGGAGTCAAGCAGTAGAGCGCTCTTCAATTTTACTGCTCGAG	3390
QY	3414	TATTTCCCTT	CTCAGATGCTGAAACGGGCAACAATTAACCTTCACTACACTTGTAG	3473
Db	3391	TACTTCTCT	CTCAGATGCTGCTGTAACGGAAACAATTAACCTTCACTACACTTGTAG	3450
QY	3474	GAA	GTCCCTTTCACAGCAGCTACCGGCACAGCCAGAGCTGGAACGGCTGATGAATCCT	3533
Db	3451	GACGTTCTT	TCACAGCAGCTACGCTCAAGCCAGAGTCTGGAACGCTCATGAATCCT	3510
QY	3534	CTCATCGACCA	TACTGTATTAACCTGAACAGAACTCAAAATCAGTCCGGAATGCCAA	3593
Db	3511	CTCATCGACCA	TACTGTATTAACCTGAACAGAACTCAAAATCAGTCCGGAATGCCAA	3570
QY	3594	AACAAGAACT	GTGTTTACGCCGTGGGCTCCAGCTGGCACTCTGTTACGCCCAAAAAC	3653
Db	3571	CAGTCAAGGCT	TCAGTTTCTCAGAGCCGGAGCGAGTGAATTCGGGACCAAGTCTTAGGAAC	3630
QY	3654	TGGCTAACCT	GGAACCCCTGTATGCGCAGCAGCGCGTTCTTAAACAACAAAACAGACAAC	3713
Db	3631	TGGCTTCC	TGGAACCCCTGTATGCGCAGCAGCGAGTATCAAGACATCTGCGATTAACAAC	3690
QY	3714	AACAGCAATTT	TACCTGCACTGTGCTTCAAAATATACTCAATGGGCGTGAATCCTC	3773
Db	3691	AACAGTGAAT	ACTCTGTGAGCTGAGCTACCAAGTACCACTCAATGGCAGAGACTCTGTG	3750
QY	3774	ATCAACCC	CTGGCACTGTATGCGCTCACACAAGACGACGAACAAGTCTTTCCATG	3833
Db	3751	GTGAATCC	GGGCCCGGCATGGCAAGCCACAAGACGATGAAGAAAGTTTTCCTCAG	3810
QY	3834	AGCGGTCTCAT	GATTTTGTGAAAAAGAGCGCCCGAGCTTCAACAACCTGCATTGGACAT	3893
Db	3811	AGCGGGGTTCT	CATCTTTTGGGAAGCAAGGCTCAGAGAAAACAAATGTGACATTGAAGAG	3870
QY	3894	GTATGATTTA	CAAGCAAGAAATTTAAAGCCACTAACCCTGTGGCCACCGAAAGATTT	3953
Db	3871	GTATGATTTA	CAAGCAAGAAATTTAAAGCCACTAACCCTGTGGCCACCGAAAGATTT	3930
QY	3954	GGAACCGTG	CGAGTCAATTTCCAGAGCAGCAGACAGACCTGCGACCGGAGATGTCAT	4013
Db	3931	GGTTCTGTAT	CTTACCAACCTCCAGAGGCAACAGACAGACGCTTACCGAGATGTCAAC	3990
QY	4014	GCTATGGAG	CATTACTGTGGCAATGTGTGGCAAGATAGACGTGTACTGACGGTCCC	4073

[illegible]



Db 1 TTGGCCACTCCCTCTCTGCGCGCTCGCTCCCTCACTGAGCGCGGCGCAACCAAGGTGCGC 60  
 QY 61 AGACGGCAGAGCTCTGCTCTGCGCGCGCGCGCGCGAGCGGCGAGAGAGGAGTG 120  
 Db 61 CGACGCCCGCGCTTTGCCCGCGCGCTCACTGAGCGAGCGGCGCAAGAGGAGTG 120  
 QY 121 GGCAACTCCATCACTAGGGGTAATCGCGAAGCGCCTCCACGCTGCCGCTCAGCGCTGA 180  
 Db 121 GCCAACTCCATCACTAGGGGT-----TCCGAGAGGGGTGAGTCTGTA 163  
 QY 181 CGTAAATTACGTATAGGG--GAGTGTCTGTATTTAGCTGTCACTGAGTGTCTTTGC 237  
 Db 164 CGTGAATTACGTATAGGGTTAGGAGGTCTGTATTAGAGGTCACTGAGTG-TTTTGC 222  
 QY 238 GACATTTTGGCACAACCACTGCGCATTTAGGGTATATATGCGCGAGTGAAGCGAGAGAT 297  
 Db 223 GACATTTTGGCACAACCATGTGTCTACGCTGGGTATTAAAGCCCGAGTGAAGCAGCAGGGT 282  
 QY 298 CTCCATTTTG-ACCGGAAATTTGAACGAGCAGCAGCCATGCCGGCTTCTAGCAGATCG 356  
 Db 283 CTCCATTTTGAAGCGGGAGTTTGAACGCGCAGCCGCGCATGCCGGGTTTACGAGATTG 342  
 QY 357 TGATCAAGGTGCGCGAGCAGCCTGAGCAGCACTGCCGGCATTTCTGACTCTGTTTGA 416  
 Db 343 TGATTAAGTCCCGACGCGCCTGACGCGCATCTGCCGCGCATTTCTGACAGCTTTGTA 402  
 QY 417 GCTGGGTGCGCGAAGGAATGGAGCTGCCCCCGGATTCTGACATGGAATCTGAATCTGA 476  
 Db 403 ACTGGGTGCGCGAAGGAATGGAGTGGAGTTGCCCGCAGATTCTGACATGGAATCTGA 462  
 QY 477 TTGAGCAGGCAACCCCTGACCGGTGCGCGAAGCTGACGCGCGACTTCTGTCCAATGCG 536  
 Db 463 TTGAGCAGGCAACCCCTGACCGGTGCGCGAAGCTGACGCGCGACTTCTGACGGAATGCG 522  
 QY 537 GCCCGGTGAGTAAGGCCCCCGGAGGCGCTTCTTTGTTGTCAGTTGAGAAAGGCGAGTCT 596  
 Db 523 GCCGTGAGTAAGGCCCCCGGAGGCGCTTCTTTGTTGTCAGTTGAGAAAGGCGAGAGCT 582  
 QY 597 ACTTCCACTCATATTTCTGTGAGAGCAAGGGGTCAAAATCATGTGTGCTGGCGCT 656  
 Db 583 ACTTCCACTCATATTTCTGTGAGAGCAAGGGGTCAAAATCATGTGTGCTGGCGCT 642  
 QY 657 TCCGTGAGTCAATTTAGGAGCAAGCTGTGAGACCATCTACCGCGGATCGAGCGGACCC 716  
 Db 643 TCCGTGAGTCAATTTAGGAGCAAGCTGTGAGACCATCTACCGCGGATCGAGCGGACCC 702  
 QY 717 TGCCCACTGTTTCCGGGTGACCAAGACCGGTATATGCGCGCGGAGGGGGAACAAGTGG 776  
 Db 703 TGCCCACTGTTTCCGGGTGACCAAGACCGGTATATGCGCGCGGAGGGGGAACAAGTGG 762  
 QY 777 TGACAGAGTGTACATCCCAACTCTCTGCGCAAGCTACGCGCGAGCTGCAAGTGG 836  
 Db 763 TGAGTGTGTCTACATCCCAACTCTCTGCGCAAGCTACGCGCGAGCTGCAAGTGG 822  
 QY 837 CGTGAATAATGAGAGATATATAGCGCTGTTGAACCTGCGCGAGCGCAACGCGC 896  
 Db 823 CGTGAATAATGAGAGATATATAGCGCTGTTGAACCTGCGCGAGCGCAACGCGC 882  
 QY 897 TCGTGCGCAGCACTGACCCACGTCAGCAGACCCAGAGAGCAACAAGAGATCTGA 956  
 Db 883 TCGTGCGCAGCACTGACCCACGTCAGCAGACCCAGAGAGCAACAAGAGATCTGA 942  
 QY 957 ACCCAATTCTGACGCGCTGTCAATCCGGTCAAAAACCTCCGCGCTACATGAGAGTG 1016  
 Db 943 ATCCCAATTCTGATGCGCGGTGATCAAAAACCTTCAGCCAGTACATGAGAGTG 1002  
 QY 1017 TCGGTGCGCTGTGAGACCGGGCATCACTCCGAGAGAGTGAATCCAGAGAGAGCCAG 1076  
 Db 1003 TCGGTGCGCTGTGAGACCGGGCATCACTCCGAGAGAGTGAATCCAGAGAGAGCCAG 1062  
 QY 1077 CCTGTAATCTCTTCAACGCGCTTCAACTGCGGTGCCAGATCAAGGCGCTCTGG 1136  
 Db 1063 CCTCATATCTCTCTCAATGCGGCTCCAACTGCGGTGCCAAATCAAGGCTGCTTGG 1122

QY 1137 ACAATGCCGCAAGATCATGCGCTGACCAAAATCCGCGCGGACTACTGTGAGCGCCG 1196  
 Db 1123 ACAATGCCGCAAGATCATGCGCTGACCAAAATCCGCGCGGACTACTGTGAGCGCCG 1182  
 QY 1197 CTCCGCGCGGACATTAACCAACCGCATCTACCGCATCTGAGCTGAACGCTACG 1256  
 Db 1183 AGCCGTGAGAGACATTTCCAGCAATCGGATTATTAATTTTGAACATAACGGTACG 1242  
 QY 1257 AACCTGCTACCGCGCTCGCTTTCTCGCTGCGCGCAAAAAGTTCCGGAAGCGCA 1316  
 Db 1243 ATCCCAATATGCGCTTCCGCTTTCTGAGATGGCCAGAAAAGTTCCGCAAGAGCA 1302  
 QY 1317 ACAATCTGCTGTTTGGCGCGCCACCAAGGCAAGCAACATCGCGGAAGCCATCG 1376  
 Db 1303 ACAATCTGCTGTTTGGCGCTGCAACTACCGGAGAGCAACATCGCGGAGCCATAG 1362  
 QY 1377 CCGACGCGCTGCGCTTCAAGCGCTGCTCACTGACCAATGAGAACTTCCCTCAATG 1436  
 Db 1363 CCGACACTGTGCGCTTCAAGCGCTGCTCACTGACCAATGAGAACTTCCCTCAATG 1422  
 QY 1437 ATTGCTGCAAGATGCTGATCTGTGAGAGAGGCAAGATGACGGCCAGGCTGG 1496  
 Db 1423 ACTGTGTCAGCAAGATGCTGATCTGTGAGAGAGGCAAGATGACGGCCAGGCTGG 1482  
 QY 1497 AGTCCGCCAAGCCATTTCTGCGCGCGCAGCAAGGTGCGCGTGAACCAAAAGTCAAGTCT 1556  
 Db 1483 AGTCCGCCAAGCCATTTCTGCGCGCGCAGCAAGGTGCGCGTGAACCAAAAGTCAAGTCT 1542  
 QY 1557 CCGCCAGATGACCCCGCGCGCTGATGCTCACTGCACTCAACCAACATGTGCGCGTGA 1616  
 Db 1543 CCGCCAGATGACCCCGCGCGCTGATGCTCACTGCACTCAACCAACATGTGCGCGTGA 1602  
 QY 1617 TTGACGGGAACAGCAACCACTTGCAGCAGCAGCCGTTGCAGAGCCGATGTTCAAT 1676  
 Db 1603 TTGACGGGAACAGCAACCACTTGCAGCAGCAGCCGTTGCAGAGCCGATGTTCAAT 1662  
 QY 1677 TTGAATCAACCGCGCTTGCAGCATGACTTTGGCAAGTGAACAAGCAGAGTCAAAAG 1736  
 Db 1663 TTGAATCAACCGCGCTTGCAGCATGACTTTGGCAAGTGAACAAGCAGAGTCAAAAG 1722  
 QY 1737 AGTTCTCCGCTGCGCGCAGATCACTGACCGAGTGGCGCATGATGTTCACTGAGAA 1796  
 Db 1723 AGTTCTCCGCTGCGCGCAGATCACTGACCGAGTGGCGCATGATGTTCACTGAGAA 1782  
 QY 1797 AGGTGAGCCCAAAAAGACCCCGCGCATGACCGCGATTAAGCGAGCCAGCGGG 1856  
 Db 1783 AGGTGAGCCCAAAAAGACCCCGCGCATGACCGCGATTAAGCGAGCCAGCGGG 1842  
 QY 1857 CCGCCCTCAGTGGCGATTCATGACGTGACGCGGAGAGAGCTCCGTTGACTTTG 1916  
 Db 1843 TCGCGAGTCACTGCGCGCATGACGTGACGCGGAGAGAGCTCCGTTGACTTTG 1899  
 QY 1917 CCGACAGGTACCAAAAATGTTCTGCTCAACCGCGCATGCTTCAAGTGTGTTCCCT 1976  
 Db 1900 CAGACAGGTACCAAAAATGTTCTGCTCAACCGCGCATGCTTCAAGTGTGTTCCCT 1959  
 QY 1977 GCAAGCATGCGAGAGATGAATCAGAATTTCAACATTTGCTTCAACGCGAGAGAG 2036  
 Db 1960 GCAAGCATGCGAGAGATGAATCAGAATTTCAACATTTGCTTCAACGCGAGAGAG 2019  
 QY 2037 ACTGTTCAAGTGTCTCCCGCGGTGTCAGAAATCTCAACCGGT--CGTCAAGAAAGAG 2093  
 Db 2020 ACTGTTCAAGTGTCTCTCC--CGTCAAGAAATCTCAACCGGT--CGTCAAGAAAGAG 2076  
 QY 2094 CGTATCGAAACTCTGTCATTCATCATCTGCTGCGCGGCTCCGAGATGCTGCT 2153  
 Db 2077 CGTATCGAAACTCTGTCATTCATCATCTGCTGCGCGGCTCCGAGATGCTGCT 2133  
 QY 2154 CCGCTGCGATCTGTCACAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 2213  
 Db 2134 CCGCTGCGATCTGTCATTCATCATCTGCTGCGCGGCTCCGAGATGCTGCT 2193



QY	2214	AAACCAAGTATGGCTGCCGATGGTTATCTTCCAGATTGGCTCGAGCAACCTCTTGAG	2273
Db	2194	AAATCAGGTATGGCTGCCGATGGTTATCTTCCAGATTGGCTCGAGCACTCTCTGAA	2253
QY	2274	GGCATTCGCGAGTGTGGCACTTGAACCTTGAGCCCGAAGCCCAAGCCAAACCAAGCA	2333
Db	2254	GGAATTAAGACAGTGTGGAAAGCTCAAACTTGCCCAACCAACCAAGCCCGCAGAGCGG	2313
QY	2334	AAGCAGGACGACGCGCGGGTCTGGTCTTCTGCTAGAGTACTCGGACCCTTCAAC	2393
Db	2314	CATAAGACGACGACGAGGGGTCTGTGCTTCTGCTAGAGTACTCGGACCCTTCAAC	2373
QY	2394	GGACTCGACAAGGGGAGCCCTCAACCGCGCGGACCGCAGCGGCCCTCGAGCAGCAAG	2453
Db	2374	GGACTCGACAAGGGAGAGCCCGTCAACGAGGACAGACCGCGGCCCTCGAGCAGCAAA	2433
QY	2454	GCCTACGACCAAGCAAGCTCAAAAGCGGGTGAACAATCCGTACTCTGCGGTATAACACGCGCAC	2513
Db	2434	GCTTACGACCGGCAAGCTCGACAGCGGAGACCAACCCGTACTCTCAAGTACCAACACGCGCAC	2493
QY	2514	GCCGAGTTTCAAGAGCGCTCTGCAAGAAGATACGTTCTTTGGGGCAACCTCGGGCGAGCA	2573
Db	2494	GCGAGTTTCAAGAGCGCTTAAAGAAGATACGTTCTTTGGGGCAACCTCGAGCGACA	2553
QY	2574	GTTCTTCAGGCGCAAGAAGCGGTTCTGCAACCTCTCGGTGCTTGAGGAAGCGCTAAG	2633
Db	2554	GTTCTTCAGGCGCAAGAAGCGGTTCTGCAACCTCTGGGCTGTTGAGGAACCTGTTAAG	2613
QY	2634	ACGGCTCCTGGAAGAAGACGTCGCGTAGAGCAGTCCGCCACAAGACCAGACTCCTCTCG	2693
Db	2614	ACGGCTCCTGGGAAAAAGAGCGCGGTAGAGCACTCTCTGTGAGCCAGACTCCTCTCG	2673
QY	2694	GGCATCGGCAAGACAGCGCCAGCACCCCGCTAAAAAGAGACTCAATTTTGGTCAGACTGGC	2753
Db	2674	GGAACCGGAAAGCGCGGCCAGCACCTTCGAAGAAAAAGATTGAATTTTGGTCAGACTGGA	2733
QY	2754	GACTCAGAGTCAGTCCCCGATTCACAACCTCTCGGAGAACCTTCAGCAACCCCGCTGCT	2813
Db	2734	GACGCAAGACTCAGTACCTGACCCCGACCTCTCGACAGCCACAGACGCCCTCTGGT	2793
QY	2814	GTTGGAACTTACTACAAATGGCTTCAGGCGGTGGCGCACCAATGGCAGACAATPAACGAAGC	2873
Db	2794	CTGGAACTAATACGATGGCTACAGGCAAGTGGCGCACCAATGGCAGACAATPAACGAGGC	2853
QY	2874	GCCGACGAGTGGGTAATGCTTCAGAAATTGGCATTTGCCAATGCATGGCTGGCGAC	2933
Db	2854	GCCGACGAGTGGGTAATTCCTCCGAAATTGGCATTTGCCAATGCATGGTGGCGAC	2913
QY	2934	AGAGTCATCACCAACGACACCGCACTGGGCTTGCCCACTACAATAACCACTCTAC	2993
Db	2914	AGAGTCATCACCAACGACACCGCACTGGGCTTGCCCACTACAATAACCACTCTAC	2973
QY	2994	AAGCAAACTCTCCAGTGTCTCAACGGGGGCCAGCAACGACAACACTACTTCGGGCTACGC	3053
Db	2974	AAACAAATTTCCAGCCAATCA--GAGGCTCGAAGCACAATCACTACTTTGGCTACAGC	3030
QY	3054	ACCCCTGGGGTATTTGATTTCAACAGATTCCAAGTCCCACTTTTCAACCAAGTACTGG	3113
Db	3031	ACCCCTGGGGTATTTGATTTCAACAGATTCCAAGTCCCACTTTTCAACCAAGTACTGG	3090
QY	3114	CAGCAGCTCATCAACAACAATTTGGGGATTTCCGGCCCAAGAGACTCAACTTCAAACTCTTC	3173
Db	3091	CAAGACTCATCAACAACAACGCGGATTTCCGACCCCAAGAGACTCAACTTCAAGCTCTTT	3150
QY	3174	AACATCCAAGTCAAGAGTCAAGCAAGATGATGGCTCAACAACCATCGCTAATAACCTT	3233
Db	3151	AACATCCAAGTCAAGAGTCAAGCAAGATGACGTAAGCAACGATTTGCCAATAACCTT	3210
QY	3234	ACCAAGACGGTTCAGAGTCTTCTCGGACTCGGAGTACCAAGCTTCCGTACGTCCTGGCTCT	3293
Db	3211	ACCAGCACGGTTCAGAGTCTTACTGACTCGGAGTACCAAGCTTCCGTACGTCCTGGCTCG	3270
QY	3294	GCGCACGAGGCTGCTCCTCCTCCGTTCCCGCGGACGCTGTCATGATTTCCGCAATACGGC	3353

Db	3271	GGCATCAAGATGCTCCCGCGTTCAGAGACGTCTTCATGCTGCGCAGATATGGA	3330
QY	3354	TACCTGACGCTCAACAATGGCAGCCAAAGCGTGGGACGTTCACTCTTTTACTGCTGAA	3413
Db	3331	TACCTACCCCTGAACAACGGGAGTCAGGACGATGAGCGCTCTTCACTTTTACTGCTGAG	3390
QY	3414	TATTTCCCTTCTCAGATGCTGAGAACGGGCAACAATTACTTCACTTACCTTTGAG	3473
Db	3391	TACTTTCCTTCTCAGATGCTGCGTACCAGAACACTTTACTTCACTTCTTTGAG	3450
QY	3474	GAAATGCTTTCCACAGACGCTACGCGCACAGCCAGACCTGGACCCTGATGAACTCT	3533
Db	3451	GACGTTCCCTTCCACAGACGCTACGCTCACAGCCAGACTGAGACCGTCTCATGAATCT	3510
QY	3534	CTCATGCACCAATACTGTATTACTCTGAACAGAACTCAAAATCAGTCCCGAAGTCCCA	3593
Db	3511	CTCATGCACCAATACTGTATTACTCTGAACAGAACTCAAAATCAGTCCCGAAGTCCCA	3570
QY	3594	AACAAGACTGTGCTTTAGCCGTGGGTCTCAGCTGCGCATGTCTGTTCAAGCCAAAAA	3653
Db	3571	CAGTCAAGCTTCAAGTTTCTCAGGCCGAGCGAGTACATTCGGACCAAGTTCAGAAC	3630
QY	3654	TGGCTACCTGGAACCTGTATTCCGACAGACGCGCTTTCTAAAACAAAAACAGAACAA	3713
Db	3631	TGGCTTCTGGAACCTGTATTCCGACAGACGCGATATCAAGACATCTGCGGATTAACA	3690
QY	3714	AACAGCAATTTTACTCTGGAAGTGTCTTCAAAATATACTCAATGGGCGGTGATCCATC	3773
Db	3691	AACAGTGAATACTGTGGAAGTGTGAGCTACCAAGTACCACTCAATGGCAGAGACTCTG	3750
QY	3774	ATCAACCCCTGGCACTGTATAGCCCTCACACAAAGACGAGAGACAAAGTCTTTCCCATG	3833
Db	3751	GTAATCCGGGCGCCCATGCGCAAGCCACAAGACGATGAAGAAAAATTTTCTCTCAG	3810
QY	3834	AGCGGTGTCAATTTTGGAAAAAGAGCGCCGAGCTTCAAACTGATGCAATGGAAT	3893
Db	3811	AGCGGGTCTCTACTTTGGGAAGCAAGGCTCAGAAAAACAATGTGAACATTTGAAAAAG	3870
QY	3894	GTCATGATTACAGACGAAGAATTAAGCCATAACCTGTGSCACCGAAAGATTT	3953
Db	3871	GTCATGATTACAGACGAAGAATTCGAAACAACCAATCCCGTGTACGAGCAAT	3930
QY	3954	GGGACCGTGGCAATTTTCCAGAGCAGACAGACCCCTGCGACCGAGATGTGCAT	4013
Db	3931	GGTCTGTATCTACCAACTCCAGAGAGGCAACAGACAGTACCGCAGATGTCAAC	3990
QY	4014	GCTATGGAGCATTACCTGGCATGTGTGGCAAGATAGAGAGTGTACCTGACGGTCCC	4073
Db	3991	ACACAAGGCGTTCTTCCAGGCATGTTGTCAGACAGAGATGTACTTTCAGGGGCC	4050
QY	4074	ATTGGGCCAAAAATCTCTCACACAGATGACACTTTCAACCCGTCTCTTATGGCGGC	4133
Db	4051	ATTGGGCCAAAAATCTCTCACACAGAGACACTTTTCAACCCCTCTCCCTCATGGGTGA	4110
QY	4134	TTTGGACTCAAGAACCCGCTCTCAGATCTCTCAATAAAAAACAGCCTGTCTCGGAAT	4193
Db	4111	TTTGGACTTAAACACCCCTCTCCACAGATTTCTCAAGAAACACCCCGTACTCGGAAT	4170
QY	4194	CCTCCGGCGAGTTTTCAGCTACAAGTTTGTCTCATTCATCACCCATATCTCCACAGA	4253
Db	4171	CCTTCGACCACTTCACTGCGGCAAGTTTGTCTTCATCACACAGTACTCCACGGGA	4230
QY	4254	CA-AGTGAAGTGAATTTGAATGGAGCTGCAAGAAAGAAAAACAGCAAGCGCTGGAATCC	4312
Db	4231	CACGGTCAAGCGTGAGATCGAGTGGAGCTGCAAGAAAGAAAAACAGCAAGCGCTGGAATCC	4290
QY	4313	CGAAGTGCAGTACATCCAAATTATGCAAAATCTGCCAACGTTATTTTACTGTGGACAA	4372
Db	4291	CGAATTCAGTACATTTCCAACTACAAACAGTCTGTTAATCGTGAATTAACCGTGAATAC	4350
QY	4373	CAATGACTTATATGAGCCTCGCCCATTTGGCAACCCGTACCTTACCCGCTCTGTA	4432

Db 4351 TAATGGCGTGTATTTCAGAGCCCTGCCCCATTTGGCACCAGATACCTGACTCGTAATCTGTA 4410  
Qy 4433 ATTACGTGTAAATCAATAAACCGGTGATTGCTTTCAGTTGAATTTGCTCTGCT 4492  
Db 4411 ATTGCTGTAAATCAATAAACCGGTGATTGCTTTCAGTTGAATTTGCTCTGCTGAT 4470  
Qy 4493 TCTTATCTTATCGGTTACCATGGTTAAGCTTACACATTAACCTGTTGGCTTGC 4552  
Db 4471 TTTCTTCTTATCTAGTTTCCATGGCTACGATAGATATATTAGCATGGCGGTTATATTA 4530  
Qy 4553 GATAAAGACTTACGTACATCGGGTTACCCCTAGTAGATGAGTTGCCACTCCCTCTG 4612  
Db 4531 CTACAAGGA-----ACCCCTAGTAGATGAGTTGGCCACTCCCTCTGCTG 4574  
Qy 4613 GCGCTCGCTCGCTCGGTGGGCGCTGCGCAACCAAGTCCGCAAGCGCAGAGCTCTGCTC 4672  
Db 4575 GCGCTCGCTCGCTCACTGAGGCGCGGCGACCAAGGTCCGCCGACGCGCGGCTTTGCC 4634  
Qy 4673 TGCCGCGCCCAACCGAGCGAGCGCGCAGAGAGGAGTGGCAA 4718  
Db 4635 GGGCGGCTCAGTAGAGCGAGCGCGCAGAGAGGAGTGGCAA 4680

RESULT 9  
LOCUS 162303 4680 bp DNA linear PAT 07-OCT-1997  
DEFINITION Sequence 1 from patent US 5658785.  
ACCESSION 162303  
VERSION 162303.1 ,GI:2480251  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 4680)  
AUTHORS Johnson, P.R.  
TITLE Adeno-associated virus materials and methods  
JOURNAL Patent: US 5658785-A 1 19-AUG-1997;  
FEATURES  
source 1. 4680  
location/Qualifiers  
BASE COUNT 1198 a 1265 c 1254 g 963 t  
ORIGIN

Query Match 65.3%; Score 3082.4; DB 6; Length 4680;  
Best Local Similarity 80.0%; Pred. No. 0;  
Matches 3781; Conservative 0; Mismatches 891; Indels 54; Gaps 11;

Qy 1 TTGCCCACTCCCTCTCTGCGCGCTCGCTCGCTGCGGCGCTGCGCAACCAAGTCCGC 60  
Db 1 TTGGCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGCGCGCGCAACCAAGTCCGC 60  
Qy 61 AGACGGCAGAGCTCTGCTGCGCGCGCCCAACCGAGCGAGCGCGCAGAGAGGAGTG 120  
Db 61 CGACGCGCGGCTTTGCGCGCGCGCTCACTGAGCGAGCGCGCAGAGAGGAGTG 120  
Qy 121 GGCACCTCCATCACTAGGGGTAATCGCGAAGCGCCTCCCAAGCTGCGCTCAAGCTGA 180  
Db 121 GCCAATCCATCACTAGGGG-----TCCCTGAGGGGTGAGTCTGA 163  
Qy 181 CGTAAATTACGTATAGG--GAGTGCTCTGTATTAGCTGTCACTGAGTCTTTGC 237  
Db 164 CGTGAATTACGTATAGGTTAGGAGGCTCTGTATTAGAGTCACTGAGTG-TTTTGC 222  
Qy 238 GACATTTTGCACACCAAGTGGCATTTAGGATATATATAGCCGAGTGAAGCAGAT 297  
Db 223 GACATTTTGCACACCAAGTGTGTACGCTGGGTATTTAAGCCCGAGTGAAGCAGAGGT 282  
Qy 298 CTCAATTG-AACGCGAATTGAAAGAGCAGCAGCATGCGGGCTTCTACGAGATG 356  
Db 283 CTCAATTGAAAGCGGAGTTGAAGCGCAGCAGCCCATGCGGGGTTTACGAGATTG 342  
Qy 357 TGATCAAGTGCAGAGCAGCTGACAGACACTGCGGGCATTTCTGACTGTTTGTGA 416  
Db 357 TGATCAAGTGCAGAGCAGCTGACAGACACTGCGGGCATTTCTGACTGTTTGTGA 416

Db 343 TGATTAAAGGTCCCGAGCGACTTGAACGGGCATCTGCGCGCATTTCTGACAGCTTTGTA 402  
Qy 417 GCTGGGTGGCCGAGAAGAAATGGAGAGCTGCCCCCGGATTTGACATGATCTGAATCTGA 476  
Db 403 ACTGGGTGGCCGAGAAGAAATGGAGAGTTGCCGCGCAGATTTGACATGATCTGAATCTGA 462  
Qy 477 TTGACAGGACACCCCTGACCGGTGGCCGAGAAGAGCTGACGCGCACTTCTGTCCAATGGC 536  
Db 463 TTGAGCAGGACACCCCTGACCGGTGGCCGAGAAGAGCTGACGCGCACTTCTGACGGAATGGC 522  
Qy 537 GCGCGGTGAGTAAGGCCCCCGAGCGCCCTCTTCTTTGTTCACTTCGAGAAGGCGAGTCT 596  
Db 523 GCGGTGTGAGTAAGGCCCCCGAGCGCCCTTTCTTTGTGCAATTTGAGAAGGAGAGAGCT 582  
Qy 597 ACTTCCACCTCCATATTCTGTTGAGAGACCAACGCGGGGTCAAATCCATGTTGCTGGCGCT 656  
Db 583 ACTTCCACATGACAGTCTGTTGAGAAACCAACGCGGGGTGAATTCATGTTTGGAGCTT 642  
Qy 657 TCTGAGTCAGATTAGGACAACTGTGTGACAGACCAATCAATCCGCGGATCGAGCCGACCC 716  
Db 643 TCTGAGTCAGATTGCGGAAAACTGATTCAAGAAATTTACCGCGGATCGAGCCGACTT 702  
Qy 717 TGCCCACTGTTTGGCGTGAACCAAGCGGTAATGGCGCGGAGGGGGAACAAGTGG 776  
Db 703 TGCCAAACTGTTTGGCGGTGCAACAAGACCAAGAAATGGCGCGGAGCGGGAACAAGTGG 762  
Qy 777 TGAACAGTGTCTACATCCCACTAATCTCTGCGCCCAAGACTCAAGCTGCAAGTGG 836  
Db 763 TGAATGAGTCTACATCCCACTAATCTCTGCTCCCAAAACCAAGCTGCAAGTGG 822  
Qy 837 CGTGAATAACATGAGGAGTATATTAAGCGGCTGTTTGAACCTGGCCGAGCGCAACGCG 896  
Db 823 CGTGAATAATATGAAACAGTATTTAAGCGGCTGTTTGAATCTCAAGAGCTTAACGCT 882  
Qy 897 TCGTGGCGCAGCAGCTGACCAAGTCAAGCAGACCAAGCAGAGCAAGAGAGATCTGA 956  
Db 883 TGGTGGCGCAGCAGCTGACCAAGTGTGTGAGAGCAGAGCAGAGCAAGAGAGATCTGA 942  
Qy 957 ACCCAATTCTGACGCGCTGTCAATCCGCTCAAAAACTCCGCGGCTACATGAGTGG 1016  
Db 943 ATCCCAATTCTGATGCGCGGTGATCAAGTCAAAAACTTCAGCCAGGTACATGAGTGG 1002  
Qy 1017 TCGGTGGCTGTGACCGGGGATCACTCCGAGAAGCAGTGGATCCAGAGACCAAG 1076  
Db 1003 TCGGTGGCTGTGACCAAGGGATTAATCTCGAGAGAGCAGTGGATCCAGAGACCAAG 1062  
Qy 1077 CCTGTACATCTCTTCAACGCGCTTCCAACTGCGGCTCCAGATCAAGCGCTTGG 1136  
Db 1063 CCTGTACATCTCTTCAATGGGCTTCCAACTGCGGCTCCAAATCAAGGCTGCTTGG 1122  
Qy 1137 ACAATGCCGCAAGATCATGGGCTGACCAAAATCCGCGGCACTAATCTGTAGGCCCG 1196  
Db 1123 ACAATGCCGCAAGATTTAGAGCTTGAATAAACCGCCGCACTAATCTGTAGGCCCG 1182  
Qy 1197 CTCCGCGCGGACATTAACCAACCGCATCTACCGCATCTGAGAGTGAAGGCTAGC 1256  
Db 1183 AGCCGCTGAGAGACATTTCCAGCAATCGATTTATAAAATTTGGAACATAACGGGTAGC 1242  
Qy 1257 AACTGCTTACGCGGCTCCGCTTTCTGCGCTGGGCGCCAGAAAAGTTCCGGAAGCGCA 1316  
Db 1243 ATCCCAATATATGCGGCTCCGCTTTCTGCGATGGGCGCAAAAAGTTCCGGAAGAGGA 1302  
Qy 1317 ACAACATCTGGCTGTTGGGCGGCGCAACCAAGGCAAGCAACATCGCGGAAGCCATG 1376  
Db 1303 ACAACATCTGGCTGTTGGGCGGCGCAACCAAGGCAAGCAACATCGCGGAAGCCATG 1362  
Qy 1377 CCGAGCGGCTGCGCTTCTACGCGCTGCGTCAATCTGAGCAATGAGAACTTCCCTTCAATG 1436  
Db 1363 CCGACACTGTGCGCTTCTACGCGGTGCGTAACTGAGCAATGAGAACTTCCCTTCAAG 1422  
Qy 1437 ATTGCTGCAACAAGATGTGATCTGTGTGAGAGAGGCGCAAGATGACGCGCAAGTCTGG 1496  
Db 1423 ACTGTGTGACAAAGATGTGATCTGTGTGAGAGAGGCGCAAGATGACGCGCAAGTCTGG 1482







Db 3631 TGGCTTCCGACCCCTGTTACCGCCAGCAGGATGATCAAGACATCTGCGGATAACAAC 3690  
Qy 3714 AACAGCAATTTTACCTGAGCTGTGCTTCAAAATATATACCTCAATGGCGGTGAATCCATC 3773  
Db 3691 AACAGTGAATACTCGTGAGCTGAGTCAACAGTACCACCTCAATGGCAGAGACTCTCTG 3750  
Qy 3774 ATCAACCTTGCACTGCTATGGCTTCAACAAGACGAGCAAGTCTTCCCATG 3833  
Db 3751 GTGAATCCGGGGCCCGCCATGCGCAAGCCACAAGACGATGAAGAAAGTTTTCCTCAG 3810  
Qy 3834 AGCGGTGTCATGATTTTGGAAAAGAGCGCGGAGCTTCAACACTGCAATTGACAT 3893  
Db 3811 AGCGGGTTCATCTTTGGGAAGCAGGCTCAGAAAAAATAATGTAACATTGAAAAG 3870  
Qy 3894 GTCATGATTAACAGACGAGAGAAATTAAGCCACTAACCTGTGGCCACCGAAGATT 3953  
Db 3871 GTCATGATTAACAGACGAGAAATCGGAACAACCAATCCCGTGGCTACGAGACGAT 3930  
Qy 3954 GGGACCGTGGCAGTCAATTTCCAGACGACGACACAGACCCCTGCGACCGGAGATG 4013  
Db 3931 GGTCTGTATCTACCAACCTCCAGAGAGGCAACAAGACGACGTAACCGCAGATGTCA 3990  
Qy 4014 GCTATGGAGCATTAACCTGCGATGTGTGGCAAGATAGAGACGTAACCTGACGGGTCC 4073  
Db 3991 ACACAGGGCTTCTTCCAGGATGTGTGGCAGACAGAGATGTACTTCAAGGGGCC 4050  
Qy 4074 ATTTGGCCAAATTCCTCACACAGATGACATTTCAACCCCTCTCTCTTATGGGGCC 4133  
Db 4051 ATCTGGCAAGATTCACACACAGGAGCAATTTCAACCCCTCTCTCTCTATGGGTGA 4110  
Qy 4134 TTTGGACTAAGAACCCGCTCTCAGATCTCATCAAAAAACGCGCTGTTCTGCGAAT 4193  
Db 4111 TTGGAATTAACACCCCTCTCAGATCTCATCAAGAACCCCGGTACTGCGAAT 4170  
Qy 4194 CTTCCGGGAGTTTTCAGCTACAAGTTTGTCTTCAATTCATCAACCAATCTCCACAGA 4253  
Db 4171 CTTGCAACCACTTCACTGCGGCAAGTTTGTCTTCAATTCATCAACAGTACTCCACGGA 4230  
Qy 4254 CA-AGTGAATGGAATTTGAATGGAGCTGACAGAAAGAAACAGCAAGCGCTGGAATCC 4312  
Db 4231 CACGGTCAAGCTGAGATCGAGTGGAGCTGACAGAAAGAAACAGCAAGCGCTGGAATCC 4290  
Qy 4313 CGAAGTGAATACACATCCAAATTAAGAAATCTGCCAAGCTGATTTTACTGTGACAA 4372  
Db 4291 CGAAATTCAGTACACTTCCAACTACAACAAGTCTGTAATCGTGAATTAACCGTGATAC 4350  
Qy 4373 CAATGACTTTATCTAGAGCTGCGCCATTTGGCACCCTTACCTTACCCGCTGTA 4432  
Db 4351 TAAATGGGTGATTCAGAGCTGCGCCATTTGGCACCAGATACCTGATCTGTAATCTGA 4410  
Qy 4433 ATTACGTTTATCAATAAACCGGTGAATTCGTTTCACTGATGAACTTGTCTCTGCT 4492  
Db 4411 ATTGCTTGAATCAATAAACCGTTAATTCGTTGAGTGAATTTGTCTCTGCTGAT 4470  
Qy 4493 TCTTATCTTATCGGTTTACCATGTTATAGCTTACATTAATCTGTTGCGCTTCCG 4552  
Db 4471 TTTCTTCTTATCTAGTTTCCATGCTAGATGATTAATTAAGCATGCGGGTTAATCA 4530  
Qy 4553 GATTAAGAATTAAGTCACTGCGGTACCCCTAGTATGATGAGTTGCGCACTCCCTCTG 4612  
Db 4531 CTACAAGG-----ACCCCTAGTATGAGTTGCGCACTCCCTCTCTG 4574  
Qy 4613 GCGTCTGCTGCTCGGTGGGCGCTGCGGACCAAGGTCCGACAGGCGAGAGCTGCTC 4672  
Db 4575 GCGTCTGCTGCTCACTGAGGCGCGGACCAAGGTGCGCCGACGCGCGGCTTGTCC 4634  
Qy 4673 TGCGGCGCCACCGAGCGAGCGCGCAGAGAGGAGTGGCAA 4718  
Db 4635 GGGCGGCTCAGTGAGCGAGCGCGCAGAGAGGAGTGGCAA 4680

RESULT 10

AX135805  
LOCUS AX135805 4675 bp DNA linear PAT 29-MAY-2001  
DEFINITION Sequence 5 from Patent WO0132711.  
ACCESSION AX135805  
VERSION AX135805.1 GI:14272040  
KEYWORDS  
SOURCE  
ORGANISM adeno-associated virus 2.  
REFERENCE adeno-associated virus 2.  
AUTHORS Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.  
TITLE 1 (bases 1 to 4675)  
JOURNAL Hermonat, P.L.  
Repo-associated virus aav rep78 major regulatory protein, mutants thereof and uses thereof  
Patent: WO 0132711-A 5 10-MAY-2001;  
THE BOARD OF TRUSTEES OF THE UNIVERSITY OF ARKANSAS (US)  
FEATURES  
source location/Qualifiers  
1..4675  
/organism="adeno-associated virus 2"  
/db\_xref="taxon:10804"  
321..2186  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAC39573.1"  
/db\_xref="GI:14272041"  
/translation="MBGFYEIVIKVPSDLGHLBGISDFVNWVAEKEMELPPDSMD  
LNLIEQPLTVAEKIQRFLTEWRVSKAPEALFFVOFEKESYFHHVLETTGVKS  
MVLGRFLSQIREKLIQRIYRIEPTLPNFAVTKRNGAGGKVVDECIENYLLPK  
TQPELOWMTNMEQYLSACINLTERKRLVAOHLTVSQTQONKONPSDAPVIRS  
KTSARMELVGNLVDKGTITSEKOWIOEDQASISFNAASNSRQIKALDNAGKIMSL  
TKTAPDYLVGQPEVDISSNRIYKILELNGYDPOYAASVFLGWAATKKFGKNTIWLFG  
PATGKTINIAEIAHTVFPYGCNVNTNENFPNDVDKVIWEEGKMTAKVESAKA  
ILGSKVRVDQKSSAQIDPTPIVITSTNMCAYIDNSTTFEHOQPLDMEKFEI  
TRRLDHFQKVTQEVKDFRMAKDHVEVEHEFFVKKGAKRPAPSADISEPKRV  
RESVAQSTSDAEASINVDARYONKCSRHVGNMMLFPCROCEARNONSNTCFTHQK  
DLECFPVSESQPVSVKKAAYOKLCYIHIMGRVDPACTACDLVNVLDLCIFEQ"

BASE COUNT 1198 a 1262 c 1251 g 964 t  
ORIGIN

Query Match 64.8%; Score 3055.8; DB 6; Length 4675;  
Best Local Similarity 79.9%; Pred. No. 0;  
Matches 3779; Conservative 0; Mismatches 887; Indels 61; Gaps 13;

Qy 1 TTGCCACTCCCTCTCTGCGGCTCGCTCGTGGGGCTCGGACCAAGTCCGC 60  
Db 1 TTGGCACTCCCTCTCTGCGGCTCGCTCGTCACTGAGCGCGGACCAAGTCCGC 60  
Qy 61 AGACGGCAGAGCTCTGCTCTGCGGCGCCCGCCACCGAGCGGAGCGGAGAGGAGTG 120  
Db 61 CGAGCGCGGCGCTTGGCGCGGCGCTCAGTGAGCGAGCGGAGCGGAGAGGAGTG 120  
Qy 121 GGCACCTCCATCAGTGGGTAATCGCGAAGCGCTCCACGCTCGCGCTCAGCGCTGA 180  
Db 121 GCGCACTCCATCAGTGGGTAATCGCGAAGCGCTCCACGCTCGCGCTCAGCGCTGA 180  
Qy 181 CGTAAATTACGTATAGG---GAGTGTCTCTGATTAAGTGTCAAGTGTCTTTCG 237  
Db 164 CGTGAATTACGTATAGGTTAGGAGGTCTCTGATTAAGTGTCAAGTGTCTTTCG 222  
Qy 238 GACATTTTGGCAGACCACTGCGCATTTAGGATATATAGCCGAGTGAGCGAGAGAT 297  
Db 223 GACATTTTGGCAGACCACTGCTGCTGAGGTATTTAAGCCCGAGTGAGCGAGAGT 282  
Qy 298 CTTCATTTTG-ACCGGAAATTGAACGAGCAGACCAATGCGCGGCTTCTACGAGATCG 356  
Db 283 CTTCATTTTGAAGCGGAGGTTTGAACGCGCAGCGCCGATGCGCGGCTTCTACGAGAT 342  
Qy 357 TGATCAAGGTGCGGAGCGCTGGAAGCAGCAGCCTGCGCGGCAATTTCTGACTCGTTGTGA 416  
Db 343 TGATTAAGGTGCGGAGCGCTTGAAGCGGAGCAGCAGCCTGCGCGGCAATTTCTGACAGCTTTGTGA 402  
Qy 417 GCTGGGTGCGGAGAGGATGGAGCTGCGCGCGGATTTCTGACATGATCTGAATCTGA 476  
Db 403 ACTGGGTGCGGAGAGGATGGAGGTTGCGCGCGGATTTCTGACATGATCTGAATCTGA 462



Db 2614 GACGGCTCCGGGAAAAAGAGCGCGGTAGAGCACTCTCTGTGAGCCAGACTCTCTC 2673  
Qy 2693 GGGCATCCGCAAGACAGAGCCAGAGCCCGCTAAAAAGAGACTCAATTTTGGTCAGACTGG 2752  
Db 2674 GGGAAACCGGAAAGCGCGGCGCAGAGCTTGCAGAAAAAGATTGATTTTGGTCAGACTGG 2733  
Qy 2753 CGACTCAGAGTCAAGTCCCCGATCCAACTCTCGAGAACTTCAGCAACCCCGCTGC 2812  
Db 2734 AGACGCAAGTCAAGTACCTGAGCCCGAGCTCTCGAGAGCCAGCAAGCCCGCTTGG 2793  
Qy 2813 TGTGGACCTACTACATGCTTCAAGCGGTGGCGCAGCAATGCGAGACAATTAACGAAG 2872  
Db 2794 TCTGGGAATAATACATGCTACAGGCAAGTGGCGCAGCAATGGCAGACAATTAACGAAG 2853  
Qy 2873 CGCCGACGAGTGGGTATGCTCAGAGAAATTGGCATTGCGATTCCACATGCTGGCGA 2932  
Db 2854 CGCCGACGAGTGGGTATGCTCAGAGAAATTGGCATTGCGATTCCACATGATGGCGA 2913  
Qy 2933 CAGAGTATCAACCAAGCAGCAGCCGCACTGGGCTTGGCCACTTAACAATACCCTCTA 2992  
Db 2914 CAGAGTATCAACCAAGCAGCAGCCGCACTGGGCTTGGCCACTTAACAATACCCTCTA 2973  
Qy 2993 CAGAGAAATCTCAGTCTTCAAGCGGGGCGCAGCAAGCAGCACTACTTGGGTACAG 3052  
Db 2974 CAAACAAATTTCCAGCCAATCA--GAGGCTCGAACGACAATCACTATTGGGTACAG 3030  
Qy 3053 CACCCCTGGGGTATTGATTCAAGATTCAGATTCAGCTTTCACACAGCTGACTG 3112  
Db 3031 CACCCCTGGGGTATTGATTCAAGATTCAGATTCAGCTTTCACACAGCTGACTG 3090  
Qy 3113 GCAAGCACTCATCAACAACAATGGGGAATTCGGCCCAAGAGACTCAACTTCAAACTTT 3172  
Db 3091 GCAAGCACTCATCAACAACAATGGGGAATTCGGCCCAAGAGACTCAACTTCAAGCTTT 3150  
Qy 3173 CAACATCCAAGTCAAGAGGTCAACAAGATGAGGTGCGTCAACAACCATGCTAATACCT 3232  
Db 3151 TAACATTCAGTCAAGAGGTCAAGAGATGAGGTGCGTCAACAACCATGCTAATACCT 3210  
Qy 3233 TACCAAGCAGGTTCAAGTCTTTCGGAATCGGAGTACAGCTTCCGTACGCTCGGCTC 3292  
Db 3211 TACCAAGCAGGTTCAAGTCTTTCGGAATCGGAGTACAGCTTCCGTACGCTCGGCTC 3270  
Qy 3293 TGGCAGCAGGGCTGCTCCCTCCGTTCCCGGCGAGCTGTTCAATGATTCGCAATACGG 3352  
Db 3271 GGGCATCAAGGATGCTCCCGCTTCCAGCAGAGCTCTTCATGCTGCGACAGTATGG 3330  
Qy 3353 CTACCTGACGCTCAACAATGGGCAAGCCGCGAGCTTCACTCTTTTACTGCTCTGA 3412  
Db 3331 ATACCTCACCTGAAACAACGGGAGTCAAGGAGTACGCTCTTACTGCTCTGA 3390  
Qy 3413 AATATTCCTTCTCAGATGCTGAGAACGGGCAACAACCTTACTCTCAGCTACACTTTGA 3472  
Db 3391 GTACTTCTCTCAGATGCTGCGTACCGGAAACAACCTTACTCTCAGCTACACTTTGA 3450  
Qy 3473 GGAAGTGCCTTTCACAGCAGTACCGGCAAGCCGAGAGCCGCTGATGAATCC 3532  
Db 3451 GGAAGTGCCTTTCACAGCAGTACCGGCAAGCCGAGAGCCGCTGATGAATCC 3510  
Qy 3533 TCTCATGCAACAATACCTGTATTAAGTGAACAAGAACTCAAGTCCGGAAGTGCCTCA 3592  
Db 3511 TCTCATGCAACAATACCTGTATTAAGTGAACAAGAACTCAAGTCCGGAAGTGCCTCA 3570  
Qy 3593 AAACAAGGACTTGTGTTAGCCGTGGGTCTCCAGCTGCGAGTGTCTGTTACGCCCAAAA 3652  
Db 3571 GCAATCAAGGCTTCAAGTTTCTCAGGCGGAGCGAGTGAACATTCGGGACCACTAGTGA 3630  
Qy 3653 CTGCTACTGAGCCCTGTTATCGGAGCAGCGCGTTTCTAAACAACAACAAGCAACA 3712  
Db 3631 CTGCTACTGAGCCCTGTTATCGGAGCAGCGCGTTTCTAAACAACAACAAGCAACA 3690  
Qy 3713 CAACGCAATTTTACTGAGCTGCTGCTTCAAAATATACTCAATGGCGGTGAATCAT 3772

Db 3691 CAACAGTGAATACCTCGTGAAGTGAAGCTTACCAAGTACCACTCAATGGCAGAGACTCT 3750  
Qy 3773 CATCAACCTTGCACTGCTATGAGCTTCAACAAGAGCAGCAAGTCTTTCCAT 3832  
Db 3751 GGTGAATC-----CGGCCATGGCAAGCCACAGAGCAGATGAAGAAAGTTTCTCTCA 3804  
Qy 3833 GAGCGGTGTCATGATTTTGAAGAAAGAGAGCGCCGAGCTTCAACACTGCATTGAGACA 3892  
Db 3805 GAGCGGGTCTCATCTTTGGGAAGCAAGGCTCAGAGAAACAATGTGAATTTGAATA 3864  
Qy 3893 TGTATGATTAACAGCAGAGAGAAATTAAGCCACTAACCTGTGGCCACCGAAGATT 3952  
Db 3865 GGTATGATTAACAGCAGAGAGAAATCGGAACAACCAATCCGTGGCTACCGAGCAGTA 3924  
Qy 3953 TGGAGCCGTGCAATTTCCAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 4012  
Db 3925 TGGTCTGTATCTACCAACTCCAGAGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 3984  
Qy 4013 TGTATGAGAGCATTACCTGAGTGTGGCAAGATGAGAGCAGTACCTGAGAGTCC 4072  
Db 3985 CACACAAGCGTCTTCCAGGAGTGTCTGAGAGCAGAGATGTACTTCAAGGCGC 4044  
Qy 4073 CATTTGGCCAAATTCCTCAACAAGATGACACTTCAACCCCTCTCTTATGGCGG 4132  
Db 4045 CATCTGGCAAGATTCACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 4104  
Qy 4133 CTTTGAATCAAGAACCCGCTCTCAGATCTCATCAAAACAAGCCTGTTCCTGCA 4192  
Db 4105 ATTGGAATTAACACCTCTCTCAAGATCTCATCAAGAACACCCCGTACTGCGAA 4164  
Qy 4193 TCTTCCGCGGAGTTTCAAGTCAAAAGTTGCTTCAATCATCAACCAATATCTCCAGG 4252  
Db 4165 TCTTCCGCGGAGTTTCAAGTCAAAAGTTGCTTCAATCATCAACCAATATCTCCAGG 4224  
Qy 4253 ACA-AGTGTGGAATGAAATGGAGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4311  
Db 4225 ACAGGCTGAGCTGAGATGAGTGGAGCTGAGAGGAGGAGGAGGAGGAGGAGGAGG 4284  
Qy 4312 CCGAAGTGCAGTACACATCCAAATATGCAAAATCTGCCAAGTTGATTTTACTGAGCA 4371  
Db 4285 CCGAATTCAGTACACTTCAACTACACAAGTCTGTAATCGTGACTTACCGTGATA 4344  
Qy 4372 ACAATGACTTATATCAGGCTCGCCCATTTGGCAGCCGCTTACCTTACCCGCTCTGT 4431  
Db 4345 CTATGCGGTGATTCAGAGCTTCCGCTTGGCAGCAGATACCTGACTCGTAATCTGT 4404  
Qy 4432 AATTAAGTGTATCAATTAACCGGTTGATTCGTTCAAGTGAACCTTGGTCTCTGTCC 4491  
Db 4405 AATGCTGTATCAATTAACCGGTTAATTCGTTCAAGTGAACCTTGGTCTCTGTCC 4464  
Qy 4492 TCTTATCTTATCGGTTACATGCTTATAGCTTATACATTAACCTGCTTGGTGGCTTGG 4551  
Db 4465 TTTCTTCTTATCTAGTTTCCATGAGCTACGTAGATTAAGCATGGCGGTTAATCATTA 4524  
Qy 4552 CGATAAAGACTTACGTCAATCGGCTTACCCCTAGTGAATGAGTTGCCACTCCCTCTCTG 4611  
Db 4525 ACTACAAGG-----ACCCCTAGTGAATGAGTTGGCCACTCCCTCTCTG 4568  
Qy 4612 CGGCTCGCTCGCTCGTGGGGCTTGGCAGCAAGTCCGAGAGCGGAGAGCTCTGCT 4671  
Db 4569 CGGCTCGCTCGCTCACTAGAGCCGGGCGAGCCAAAGTCCGCCAGCGCCGCTTGGCC 4628  
Qy 4672 CTGCGGCGCCCAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 4718  
Db 4629 CGGCGGCGCTCAGTGAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 4675

RESULT 11  
AX286292 4675 bp DNA linear PAT 21-NOV-2001  
LOCUS AX286292  
DEFINITION Sequence 1 from Patent WO0180840.  
ACCESSION AX286292  
VERSION AX286292.1 GI:17048540



	KEYWORDS	:	adeno-associated virus 2.
SOURCE	: adeno-associated virus 2		
ORGANISM	: viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.		
REFERENCE .	1		Raj, K. and Beard, P.M..
AUTHORS			Cytotoxic agents
TITLE			Patent: WO 0180840-A 1 01-NOV-2001;
JOURNAL			BIG INTERNATIONAL LIMITED (GB)
FEATURES			location/Qualifiers
source		1..4675	/organism="adeno-associated virus 2"
			/db_xref="taxon:10804"
BASE COUNT		1198 a	1262 c
ORIGIN			1251 g
			964 t
Query Match		64.8%; Score 3055.8;	DB 6; length 4675;
Best Local Similarity		79.9%;	Pred. No. 0;
Matches 3779;	Conservative	0;	Mismatches 887; Indels 61; Gaps 13
QY	1 TTGCGCACTCCCTCTCTGTGGCGCCTGCCTTCGTTGGTGGGCCCTGCCAGCAAAGTCCGC	60	
Dd	1 TTGGCCAATCCCCTCTCTGTGGCGCCTGCCTTCGTTACTTAGTGCCGGCACCAAGTCCGC	60	
QY	61 AGACGGCAGAAGCTCTGCTCTGGCCCCACCAGCGAGCGAAGCGGAGGAATTG	120	
Dd	61 CGACGCCCGGGCTTTGGCCCGGGCGGCTTAAGCGAGCGAGCGGCGAGAGGAAAGT	120	
QY	121 GGCACTCATCACTAGGGGTAATCGGAAGCGCCTCCACGCTGCCGCTACGCGCTGA	180	
Dd	121 GCCAACCTCATCACTAGGGG-----TCCTGAGGGGTGAGTCTGTA	163	
QY	181 CATAATTACGTATAGG---GAGTGTCTGTATTAGCTGTACAGTGAAGTCTTTGC	237	
Dd	164 CGTGAATTACGTATAGGTTAGGAGGTCTGTATTAGAGTCAAGTGAAG-TTTTTG	222	
QY	238 GACAATTTGCGACACCAACGAGTGGCCATTAGGATATATAGGCCGAGTGACGACAGAT	297	
Dd	223 GACAATTTGCGACACCAATGTGTCAACGCTGGTATTTAAGCCCGAGTGACACGACGGT	282	
QY	298 CTCCATTTTG-AACCGCAAAATTGAACGAGCAGCAGCCATGCCGGCTTTCAAGAGATCG	356	
Dd	283 CTCATTTTGAAGCGGAGGTGGAACGCGCAGCCGCATGCCGGGTTTTCAGAGATTG	342	
QY	357 TGATCAAGGTGCCGAGCGACCTGAGCAGCACTGCCGGCATTCTGACTGTTGTGA	416	
Dd	343 TGAATTAAGTCCCCAGCACCTTGAACGGGCAATCTGCCGCATTCTGACAGCTTTGTGA	402	
QY	417 GCTGGGTGCCGAGAAGGAATGGAGCTGCCCGGATTTCTGACATGAATCTGA	476	
Dd	403 ACTGGGTGCCGAGAAGGAATGGAGTGTGCCGCAGATTTCTGACATGAATCTGA	462	
QY	477 TTGAGCAGGACCCCCGTGACCGGTGCCGAGAAGCTGCAGCGGCACTTCGTGCCAATGGC	536	
Dd	463 TTGAGCAGGACCCCCGTGACCGGTGCCGAGAAGCTGCAGCGGCACTTCGTGCCAATGGC	522	
QY	537 GCCGCGTGAAGAGGCCCGGAGGCCCTCTTTTGTTCAGTTCGAGAAGGGCGAGTCT	596	
Dd	523 GCCGCGTGAAGAGGCCCGGAGGCCCTTTTCTTTGTGCAATTGAGAAGGGAGAGAGCT	582	
QY	597 ACITTCACCTCATATTTCTGTGTGAGACCAAGGGGGTCAATTCATGTTGTGGCCGCT	656	
Dd	583 ACTTCCACATGACAGTGTCTGTGAACCAAGGGGGTGAATTCATGTTTGGAGCTT	642	
QY	657 TCCTGAGTCAGATTAGGAGCAAGCTGTGAGACCATCTACCGCGGATCGAGCCGACCC	716	
Dd	643 TCCTGAGTCAGATTTCGGAAAACTGATTCAGAGAAATTAACGCGGATCGAGCCGACTT	702	
QY	717 TGCCCAACTGTTTCGGGTGACCAAGACGGTTAATGGCGCCGAGGGGGAACAAGTGG	776	
Dd	703 TGCCAAAAGTGTTCGGGTGACCAAAAGACCAAAATGGCGCCGAGGGGGAACAAGTGG	762	
QY	777 TGAGCAGTGTCAATCCCCCAACTACTCTCTGCCCCAAGACTAGGCCGAGCTGCAGTGGG	836	

Db	763	TGATGAGTGTACATCCCCAATTACTTGTCTCCCAAAACCAGCCTGAGCTCCAGTGGG	822
Qy	837	CGTGACTAACATGAGAGGATATATAAGCCGCTGTTGAACCTGCGCGCAACCGC	896
Db	823	CGTGACTAATATGGAACAGTATTTAAGCGCTGTTGAATCTCACGGAGGTAAACGGT	882
Qy	897	TCGTGGCGCAGCACTGACCCACGTCAGCCAGACCAGAGCAGAAACAAGAAATCTGA	956
Db	883	TGCTGGCGCAGCATCTGACGCACGTGTCCGACAGCAGAGCAGAAACAAGAAATCAGA	942
Qy	957	ACCCCAATTCTGACGCGCCTGTCTATCCGGTCAAAAACCTCCGCGCTACATGAGCTGG	1016
Db	943	ATCCCAATTCTGATGCGCCGGTGATCAGATCAAAAACCTTCAGCCAGGTACATGAGCTGG	1002
Qy	1017	TCGGGTGGCTGTGGAACCGGGGATCACTCCGAGAAAGCATGTGATCCAGAGGACCAGG	1076
Db	1003	TCGGGTGGCTGTGGAACAAGGGGATTAACCTCGGAGAAAGCATGTGATCCAGAGGACCAGG	1062
Qy	1077	CCTCGTACATCTCTTCAACGCGCCTTCCAACTGCGCGTCCAGATCAAGCGCTCTGG	1136
Db	1063	CCTCATACATCTCTTCAATGCGCGCTCCAACTGCGCGTCCCAATCAAGCGCTCTGG	1122
Qy	1137	ACAATGCCGCGCAAGATCATGCGCTGACCAAAATCCGCGCCGACTACTGTAGGCCCGG	1196
Db	1123	ACAATGCCGCGCAAGATATAGAGCTGACTAAACCGCCCGACTACTGTGTGGCCAGC	1182
Qy	1197	CTCCGCGCGCGACATTAACAACCAACCGCATCTACCGATCTGAGCTGAACGGCTACG	1256
Db	1183	AGCCCGTGAGGACATTTCCAGCAATCCGATTTATAAATTTTGAACCTAAACGGGTACG	1242
Qy	1257	AACTGCTACGCGCGCTCCGCTCTTCTCGGCTGGGCCAGAAAAGGTTCCGGAAAGCGCA	1316
Db	1243	ATCCCAATATGCGGCTTCCGCTTCTTGTGGATGGGCCACGAAAAGTTCCGCAAGAGA	1302
Qy	1317	ACAACCATGTGGCTGTTGGGCGCGGCCAACCAACCGGCAAGAACCAATCCGCGAAGCCATCG	1376
Db	1303	ACAACCATGTGGCTGTTGGGCGCTGCAACTACCGGGAAGAACCAATCCGCGAGGCCATAG	1362
Qy	1377	CCCAACGCGTGCCCTTCTACGCGCTGCTCACTGGAACCAATGAACTTCCCTTCAATG	1436
Db	1363	CCCACTGTGCGCTTCTACGCGGTGCGTAACTGGAACCAATGAACTTCCCTTCAACG	1422
Qy	1437	ATTGCTGCAACAAGATGTGATCTGTGTGGAGAGGGCAAGATGACCGCCAAAGTCTGG	1496
Db	1423	ACTGTGTGCAACAAGATGTGATCTGTGTGGAGAGGGGAAGATGACCGCCAAAGTCTGG	1482
Qy	1497	AGTCCGCCAAGGCCATTTCTCGCGCGCAGCAAGTGCCTGTGACCAAAAAGTGAAGTCTG	1556
Db	1483	AGTCCGCCAAGGCCATTTCTCGGAGGAAGCAAGTGCCTGTGACCAAAAAGTGAAGTCTG	1542
Qy	1557	CCGCCCCAGATGACCCCCACCCCGTGATGCTACCTCCCAACCAACATGTGCGCCGTGA	1616
Db	1543	CGGCCCCAGATGACCCCCACCTCGTGATGCTACCTCCCAACCAACATGTGCGCCGTGA	1602
Qy	1617	TTGACGGGAACAGCACCACTTCGAGCACACGAGCCGTTGACAGGACCCGATGTTCAAT	1676
Db	1603	TTGACGGGAACACTCAAGACCTTCGAACACACGAGCCGTTGCAAGACCCGATGTTCAAT	1662
Qy	1677	TTGAACTCAACCCGCGTCTGAGCATGACTTTGGCAAGGTGACAAAGCAGGAATCAAG	1736
Db	1663	TTGAACTCAACCCGCGTCTGAGCATGACTTTGGGAAGTCAACAAAGCAGGAATCAAG	1722
Qy	1737	AGTTCTTCGCTGGCGCAGGATCAAGTACCGAGGTGGCGCATGAGTTCTACGTCAAG	1796
Db	1723	ACTTTTTCGCTGGCGCAAGGATCAAGTGTGAGGTGAAGCATGAATTTCTACGTCAAAA	1782
Qy	1797	AGGGTGAAGCCAACAAAGACCCCGCCCGATGACGCGGATTAAGCGAAGCCCAAGCGGG	1856
Db	1783	AGGGTGAAGCCAAGAAAGACCCCGCCCGATGACGCGGATTAAGTGAAGCCCAAGCGGG	1842
Qy	1857	CCTGCCCTCAGTCCGGGATTCATGACGTCAAGCGCGGAAGAGCTCCGGTGGACTTTG	1916

Db 1843 TGGCGAGTCAGTTGCGCAGCCATCGAGCTCAGACGCGGA--AGCTTCGATCAACTACG 1899  
Qy 1917 CCGACAGGTACCAAAACAATGTTCTCGTACGCGGGCATGCTTCAGATGCTGTTCCCT 1976  
Db 1900 CAGACAGGTACCAAAACAATGTTCTCGTACGCGGGCATGATGATGCTGTTCCCT 1959  
Qy 1977 GCAAGACATGCGAGAGATGAATCAGAAATTCACATTTGCTTCACGACGCGAGAGAG 2036  
Db 1960 GCAGACAAATGCGAGAGATGAATCAGAAATTCAAATATCTGCTTCACACGACAGAGAG 2019  
Qy 2037 ACTGTTCAAGATGCTTCCCGCGGTTCAGAAATCTCAACCCGT---CGTCAGAAAGAGA 2093  
Db 2020 ACTGTTAGAGTGCTTCC--CGTTCAGAAATCTCAACCCGTTTCTGTCGCAAAAAGG 2076  
Qy 2094 CGTATCGGAAATCTCTGTGCCATTCATCATCTGCTGGGCGGCTCCGAGATTGCTGCT 2153  
Db 2077 CGTATCAGAAACTGTGCTACATTCATCATATCATGGAAGGTGCCAGA--CGCTTGCA 2133  
Qy 2154 CCGCTGCGATCTGTCAACGTGACCTGATGACTGTGTTCTGAGCAATAATGACTT 2213  
Db 2134 CTGCTGCGATCTGTCAATGTGATTTGATGACTGATCTGATCTTTGAACAATAATGATTT 2193  
Qy 2214 AAACCAAGTATGCTGCGAGTGTATCTTCCAGATTGCTCGAGCAACCTCTCTGAG 2273  
Db 2194 AAATCAGGTATGCTGCGAGTGTATCTTCCAGATTGCTCGAGCAACTCTCTGAA 2253  
Qy 2274 GGCATTCGCGAGTGTGGACTTGAAACTTGAGCCCCGGAAGCCCAAGCCCAACGCA 2333  
Db 2254 GGAATAGACAGTGTGTGAAGCTCAAACTTGCCCCACCAACCAAGCCCGCAGAGCGG 2313  
Qy 2334 AAGCAGACGACGCGCCGGGTCTGTGCTTCTGCTGCTACAGTACCTCGGACCTTCAAC 2393  
Db 2314 CATAGGACGACGACGAGGGGTCTGTGCTTCTGCTGCTACAGTACCTCGGACCTTCAAC 2373  
Qy 2394 GGAATCGACAGAGGGGAGCCCGTCAACGCGGCGGACGACGCGCTCGAGCAG-ACAA 2452  
Db 2374 GGAATCGACAGAGGGAGCCCGTCAACGAGGACGACGCGCTCGAGCAGCTACAA 2433  
Qy 2453 GGCCTACGACGACGCTCAAAAGCGGCTGACATCCGTACCTGCGTATTAACGACGCGA 2512  
Db 2434 AGCCTACGACGCGCAGCTCGACAGCGGACAAACCGTACCTCAAGTACAAACGCGCA 2493  
Qy 2513 CGCGAGTTTCAAGAGCGTCTGCAAGAGATACGTCTTTGGGGCAACCTCGGCGAGC 2572  
Db 2494 CGCGAGTTTCAAGAGCGCTTAAGAGATACGTCTTTGGGGCAACCTCGGACGAGC 2553  
Qy 2573 AGTCTTCAAGCGCCAAAGAGCGGTTCTGCAACTCTCGGTCTGTTGAGGAAGCGCTAA 2632  
Db 2554 AGTCTTCAAGCGCCAAAGAGCGGTTCTGCAACTCTCGGTCTGTTGAGGAACCTGTTAA 2613  
Qy 2633 GACGCTCTCTGAAAAGAAAGCTCCGCTAGAGCAGTCCGACCAAGAGCCAGACTCTCTC 2692  
Db 2614 GACGCTCTCGGAAAAAAGAGCGCGTACACTCTCTGTGAGCCAGACTCTCTCTC 2673  
Qy 2693 GGGCATCGGCAAGACAGGCCAGCGCCGCTAAAGAGAGACTCAATTTGTCAGACTGG 2752  
Db 2674 GGGAAACGGAAAGCGCGGCGACGAGCTGCAAGAAAAAGATTGATTTGTCAGACTGG 2733  
Qy 2753 CGACTCAGATCAGTCCCGCATCCACACTCTCGAGAACCTCAGCAACCCCGCTGC 2812  
Db 2734 AGACGCACTCAGTACCTGACCCCGAGCTCTCGACAGCCACGAGCCCGCTCTGG 2793  
Qy 2813 TGTGGACCTACTACATGCTTCAAGCGGTGGCGGACCAATGGCAGACATAAGAGAG 2872  
Db 2794 TCTGGGAACTAATAAGATGCTACAGGCAATGGCGCACCAATGGCAGACATAAGAGAG 2853  
Qy 2873 CGCCGACGAGTGGGTATGCTCAGGAAATTTGGCATTTGCGATTCCATGCTGGGCGA 2932  
Db 2854 CGCCGACGAGTGGGTATTTCTCCGAAATTTGGCATTTGCGATTCCATGATGGGCGA 2913  
Qy 2933 CAGAGTCAATCAACCAAGCAGACCCGCACTGGGCTTGGCCACCTAATTAACCACTCTTA 2992  
Db 2914 CAGAGTCAATCAACCAAGCAGACCCGCACTGGGCTTGGCCACCTAATTAACCACTCTTA 2973

Qy 2993 CAACGAATTCAGTGTCTTCAACGGGGCCAGCAACCAACCACTACTTCGGCTACAG 3052  
Db 2974 CAACGAATTCAGGCAATCA--GGAGCCTCGAACCAATCACTACTTTGGCTACAG 3030  
Qy 3053 CACCCCTGGGGTATTTGATTTCAACAGATTCCACTGCCACTTTTACCGACGTAAGT 3112  
Db 3031 CACCCCTGGGGTATTTGATTTCAACAGATTCCACTGCCACTTTTACCGACGTAAGT 3090  
Qy 3113 GCAGGCACTCATCAACGAATTTGGGATTTCCGGCCCAAGAGACTCAACTTCAACTCTT 3172  
Db 3091 GCAAGACTCATCAACGAATTTGGGATTTCCGACCGCAAGAGACTCAACTTCAAGCTCTT 3150  
Qy 3173 CAACATCCAAGTCAAGAGGTCAAGAGATGATGCGTCAACCACTCGCTAATAACT 3232  
Db 3151 TAACATTCAGTCAAGAGGTCAAGAGATGATGCGTCAACCACTCGCTAATAACT 3210  
Qy 3233 TACGACGAGTTCAGTCTTCTCGGACTCGGATACCACTTCCGTACGCTCGGCTC 3292  
Db 3211 TACGACGAGTTCAGTCTTCTCGGACTCGGATACCACTTCCGTACGCTCGGCTC 3270  
Qy 3293 TGGGCAACGAGGTGCTGCTCCCTCGTTCGCGGCGAGTGTATGATTTCCGCAATACG 3352  
Db 3271 GGGCATCAAGATGCTTCCCGCTCGTTCGCGGCGAGTGTATGATTTCCGCAATACG 3330  
Qy 3353 CTACCTGACGCTCAACGAATGCGACCAAGCCGTGGAGCTCATCTTTTACTGCTGGA 3412  
Db 3331 ATACCTCAACCTGAAACAGGAGTCAAGCAGTACGCTCTTCAATTTACTGCTGGA 3390  
Qy 3413 ATATTTCCCTTCTCAGATGCTGAAGACGGCAACCACTTACCTTCACTACACTTGA 3472  
Db 3391 GATCTTCTCTCAGATGCTGCGTACCGGAAACCACTTACCTTCACTACACTTGA 3450  
Qy 3473 GGAAGTGCCTTTCCACAGAGCTACCGGACAGCCAGCCCTGAGCCGCTGATGAATCC 3532  
Db 3451 GGAAGTGCCTTTCCACAGAGCTACCGTACAGCCAGCTGAGCCGCTGATGAATCC 3510  
Qy 3533 TCTCATGCAACCAATACCTGATTAACCTGAACGAATCAAAATCAGTCCGGAAGTCCCA 3592  
Db 3511 TCTCATGCAACCAATACCTGATTAACCTGAACGAATCAAAATCAGTCCGGAAGTCCCA 3570  
Qy 3593 AAACAGGACTTGTGTTTAAAGCGGTGCTCCAGCTGCGATGTCTGTTACGCCCAAAA 3652  
Db 3571 GCAATCAAGCTTCAAGTTTCTCAGCGCGGAGCGAGTGAATTCGGGACGATTAAGAA 3630  
Qy 3653 CTGGCTACCTGACCTGTTATTCGCGAGCAGCGCTTCTTAAACAAAAACAGACAA 3712  
Db 3631 CTGGCTTCTGACCTGTTATTCGCGAGCAGCGCTTCTTAAACAAAAACAGACAA 3690  
Qy 3713 CAACAGCAATTTTACCTGAGCTGCTTCAAAATTAACCTCAATGGGCTGAATCCAT 3772  
Db 3691 CAACAGCAATTTTACCTGAGCTGCTTCAAAATTAACCTCAATGGGCTGAATCCAT 3750  
Qy 3773 CATCAACCTTGCACTGCTATGCGCTCACACAAAGAGCAGAAACAAGTTCTTCCAT 3832  
Db 3751 GGTGAATC-----CGGCAATGGCAAGCCACAAAGAGCAGTGAAGAAAGTTTCTCTCA 3804  
Qy 3833 GAGCGGTGTCATGATTTTGGAAAAAGAGCGCCGAGCTTCAAAACATGCAATGGACAA 3892  
Db 3805 GAGCGGTGTCATGATTTTGGAAAAAGAGCGCCGAGCTTCAAAACATGCAATGGACAA 3864  
Qy 3893 TGTATGATTTACAGAGAGAAATTAAGCCACTAACCTGTGGCCACCGAAGATT 3952  
Db 3865 GGTATGATTTACAGAGAGAAATTAAGCCACTAACCTGTGGCCACCGAAGATT 3924  
Qy 3953 TGGGACGTTGAGTCAATTTCCAGAGCAGCAGACAGCCCTGCGACCGAGATGCA 4012  
Db 3925 TGGTCTGTATTAACCACTCCAGAGAGCAGCAGCAGCAAGCAAGTACCGCAGATGCA 3984  
Qy 4013 TGTATGAGGATTAACCTGCGATGCTGTGGCAAGTAGAGAGTACCTGAGGCTC 4072  
Db 3985 CACACAGGCGTCTTCCAGGCAATGCTGTGGCAGGACAGAGATGTGTACTTCAAGGCGC 4044

QY	4073	CATTGGGGCCAAATTCCTCTACACAGATGGACACTTTCACCCGCTCTCTTATGGCGG	4132
Db	4045	CATCTGGGCAAAGATTCCACACGGAACGGACATTTTCAACCCTCTCCCTCATGGGTGG	4104
QY	4133	CTTTGGACTCAAGAACCCGCGCTCTCAGATCTCTCATCAAAAAACAGCCTGTTCTCGGAA	4192
Db	4105	ATTGGGACTTAACACCCCTCTCTCCACAGATTCTCATCAAGAACACCCCGGTACCTGCGAA	4164
QY	4193	TCCTCCGGCGGAGTTTTCAGCTACAAAGTTTGCTTCAITTCATCACCCAATACTCCACAG	4252
Db	4165	TCCTTCGACCACCTTCAGTGGCGCAAGTTGCTTCTTCATCAACACAGTACTCCACGGG	4224
QY	4253	ACA-AGTAGTGTGMAATTGAATGGGAGCTGCAGAAAGAAAACAGCAAGCGCTGGAATC	4311
Db	4225	ACACGGTCAGCGTGAGATCGATGGGAGCTGCAGAAAGAAAACAGCAACGCTGGAATC	4284
QY	4312	CCGAAGTCAGTACACATCCAAATTATGCAAAATCTGCCAACGTTGATTTTACTGTGACA	4371
Db	4285	CCGAATTCAGTACACTTCCAACTAACAAAGTCTGTTAATCGTGAATTAACCGTGATA	4344
QY	4372	ACAATGGA,CTTATACTGAGCCTCGCCCATTTGGACACCGCTTACCTTACCCGTCCTGT	4431
Db	4345	CTAATGGCGGTGATTCAGAGCCTCGCCCATTTGGACACCAAGATACTGACTCGTAATCTGT	4404
QY	4432	AATTACGTGTTAATCAATAAACCGGTTGATTCGTTTCAGTTGAACTTTGGTCTCCTGTCC	4491
Db	4405	AATTGCTTGTATATCAATAAACCGTTTAATTCGTTTCAGTTGAACTTTGGTCTCCTGTCC	4464
QY	4492	TTCTTATCTTATCGGTTACCATGGTTTATAGCTTACACATTA,CTGCTTGGTTCGCTTCG	4551
Db	4465	TTTCTTCTTATCTAGTTCATGGCTACGTAGATAAGTAGCATGGCGGTTAATCATTA	4524
QY	4552	CGATAAAGACTTACGTCATCGGGTTACCCCTAGTAGTAGGTTGCCACTCCCTCTCTG	4611
Db	4525	ACTACAAGGA-----ACCCCTAGTAGTAGGTTGGCCACTCCCTCTCTG	4568
QY	4612	CGCGCTCGCTCGCTCGGTGGGCTCGCGAACCAAAAGTCCCGACAGACGGCAGAGCTTGCT	4671
Db	4569	CGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAAGTGGCCCGACGCCCGGGCTTGGCC	4628
QY	4672	CTGCCGGGCCCCACCGAGCGAGCGAGCGCGCAGAGAGGGAGTGGGCAA	4718
Db	4629	CGGGCGGCTCAGTAGCGGAGCGAGCGCGCAGAGAGGGAGTGGGCAA	4675

AA2CG	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE	PUBMED	REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE	PUBMED	FEATURES	SOURCE		
AA2CG		Adeno-associated virus 2, complete genome.	J01901 M12405 M12468 M12469	J01901.1 GI:209616	alternative splicing; complete genome; major coat protein.	Adeno-associated virus 2 DNA from human HeLa cells.	Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.	1 (bases 4532 to 4675)	Samulski,R.J., Sriavastava,A., Berns,K.I. and Muzyczka,N.	Rescue of adeno-associated virus from recombinant plasmids: gene correction within the terminal repeats of AMV	Cell 33 (Y), 135-143 (1983)	84282662	6088052	2 (bases 1 to 4675)	Sriavastava,A., Luby,E.W. and Berns,K.I.	Nucleotide sequence and organization of the adeno-associated virus 2 genome	J. Virol. 45 (2), 555-564 (1983)	83164299	6300419		Location/Qualifiers	1..4675	/organism="adeno-associated virus 2"

```

repeat_region      1. .145
                   /note="5' inverted terminal repeat"
misc_feature       42. .83
                   /note="flip oriented DNA"
mRNA               287. .4447
                   /note="major coat protein A mRNA (alt.)"
                   join(321. .1906,2228. .2252)
                   /note="major coat protein A' (alt.)"
                   /codon_start=1
                   /protein_id="AAA42372.1"
                   /db_xref="GI:209617"
                   /translation="MPGFYEIVIKVPSDLDGHLPGISDSFVNWVAEKEMELPPDSOMD
LNLIEOAPLTVAEKLRDPLEFTEMRWRSKAPFALPFVQFEKGESEFHMHLVETTGVS
MVLGRFLSQIREKLIQRIYFGLIEPTLPWFVAVTYKTRNGAGKNVDECIIPNYLLPK
TQELQWAMTNEOYLVSACLNLTERKRLVAQHLFVHSQIQEONKENQNPNSDAPVIRS
KTSARYMELVGMVLDKGITSEKOWIQEDQASYISFNASNSRSQIKALDNAGKIMSL
TKTAPDYLVGQDPVEDISSNRIYKILLENGYDPQYAAVFLGWAATKKFGKRNITWLF
PATGKTNIABALAHTVPFYGCVMNTNENFPNDQCDKWIMWEEGKMTAKVESAKA
ILGGSKVRVDQKCKSSAQIDPTPVIIVTSNTNMCAVLDGNSSTFEHQPLQDRMFKFEL
TRRLDHDGFKVTKQEKVDFFRWAKDHVEVEHEFFYVKKGGAKKRPAPSDADISEPKRV
RESVAQPTSDAEASINVADRLARGHSL"
CDS                321. .2186

```

/note="major coat protein A"  
 /codon\_start=1  
 /protein\_id="AA42374.1"  
 /db\_xref="GI:209619"  
 /translation="MPGEYEVIVKPSDL DGHLPGISDSFVNWVAEKEMELPPDSMD  
 LNLIEQAPLTVAEKLRQDFLTWRVRSKAPALFVDFEKGESYFHMVLVETTGVS  
 MVLGRFLSQIREKLIQRIYRGIEPTLPNMFVATKTRNGAGGNKVDECIYNYLLPK  
 TOPELQWAWTMEQYLISACINLTERRLVAQHLTHVSTOEONKONPNSDAPVIRS  
 KTSARYMELVGMVLDKGI TSEKOWIOEDQASYISFNAANSRSQIKALDNAGKIMSL  
 TKTAPDYLVGQOPVEDISSNRIYKILELNGYDPQYASVFLGWATKKFKRNTIWLFG  
 PATGKTNIAEIAHTVPFYGCNVNTNENPFNDCKMVIWEEGKMTAKVESAKA  
 ILGSKVRVDQCKSSAQIDPTPVI VTSNTMCAVIDGNSTTFEHQPLQDRMFKEFL  
 TRRLDHDGKVTQKEVKDFFRWAKDHVEVEHEFYKKGAKKRPAPSDADISEPKRV  
 RESVAOPSTSDAEASINYADRYONKCSRHVGMNMLFPQRQERMNQNSNICFTHGOK  
 DCECFPVSESQPVSVKKAYQKLCYIHIMGKVPACTACDLVNVLDLDCIFEEQ"  
 873.4447  
 /note="major coat protein A mRNA (alt.)"  
 join(993.1906,2228.2252)  
 /note="major coat protein Aa (alt.)"  
 /codon\_start=1  
 /protein\_id="AA42373.1"  
 /db\_xref="GI:209618"  
 /translation="MELVGMVLDKGI TSEKOWIOEDQASYISFNAANSRSQIKALD  
 NAGKIMSLTKTAPDYLVGQOPVEDISSNRIYKILELNGYDPQYASVFLGWATKKFGK  
 RNTIWLFGPATGKTNIAEIAHTVPFYGCNVNTNENPFNDCKMVIWEEGKMTA  
 KVESAKAILGSKVRVDQCKSSAQIDPTPVI VTSNTMCAVIDGNSTTFEHQPLQ  
 DRMFKEFLTRRLDHDGKVTQKEVKDFFRWAKDHVEVEHEFYKKGAKKRPAPSDA  
 DISEPKRVRESVAOPSTSDAEASINYADRLARGHSL"  
 993.2186  
 /note="major coat protein A' (alt.)"  
 /codon\_start=1  
 /protein\_id="AA42375.1"  
 /db\_xref="GI:209620"  
 /translation="MELVGMVLDKGI TSEKOWIOEDQASYISFNAANSRSQIKALD  
 NAGKIMSLTKTAPDYLVGQOPVEDISSNRIYKILELNGYDPQYASVFLGWATKKFGK  
 RNTIWLFGPATGKTNIAEIAHTVPFYGCNVNTNENPFNDCKMVIWEEGKMTA  
 KVESAKAILGSKVRVDQCKSSAQIDPTPVI VTSNTMCAVIDGNSTTFEHQPLQ  
 DRMFKEFLTRRLDHDGKVTQKEVKDFFRWAKDHVEVEHEFYKKGAKKRPAPSDA  
 DISEPKRVRESVAOPSTSDAEASINYADRYONKCSRHVGMNMLFPQRQERMNQNSN  
 ICFTHGOKDCECFPVSESQPVSVKKAYQKLCYIHIMGKVPACTACDLVNVLDLDCI  
 FEEQ"  
 1853.4447  
 /note="major coat protein B mRNA (alt.)"  
 1907.2227  
 /note="major coat protein A intron"  
 2810.4324  
 /note="major coat protein B"  
 /codon\_start=1  
 /protein\_id="AA42376.1"  
 /db\_xref="GI:209621"  
 /translation="MATGSGAPMADNNEGADVGNSGNWHCDSTWMDRVITTSIRIT



WALPTYNHLYKQISSQSGASNDNHFGYSTPWGYPEDNRFHCHFSPRDWORLNNMW  
GFRPKRLNFKLFNIQVKEVTQNDGTTTIANNLSTVOVFTDSEYQLPYVLGSAHQGL  
PPFPADVEMPOYGYLTNLNGSOYVGRSSFYCLEYFSPOMLRGNNFTFSPYEDVPE  
HSSYAHQSOLDRLMPLIDYLYLWLSRTNTPSGTTQSRLOFSQAGASDIRDSRNL  
GPGCYROORVSKTSADNNNSSEYSGATKYLNGRSLVNPAMASHKDEKPEPQSG  
VLIFGKQSEKTNVNIKYMITDEEIGTTNPVATEOYGSVSTNLQRNRQAATADV  
TQGLVPGMWQDRDVLQGLIWKIPIHTDGHFHPSPLMGFGKHPPOILIKNTPVP  
ANPSTFSAAKFASFITQYSTGHQRGDRVGAEGKQTLLESRNSVHFQLOQVC"  
repeat\_region 4531.4675  
/note="3' inverted terminal repeat"  
misc\_feature 4592.4634  
/note="flop oriented DNA"  
BASE COUNT 1198 a 1262 c 1251 g 964 t  
ORIGIN 5' end of genomic DNA.

Query Match 64.8%; Score 3055.8; DB 14; Length 4675;  
Best Local Similarity 79.9%; Pred. No. 0;  
Matches 3779; Conservative 0; Mismatches 887; Indels 61; Gaps 13;

QY 1 TTGCCCACTCCCTCTCTGCGCGCTCGCTCGCTCGGTGGGCGCTGGCAACCAAGTCCGC 60  
Db 1 TTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCGCGGCGCAACCAAGTCCGC 60  
QY 61 AGACGGCAGAGCTCTGCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120  
Db 61 CGACGCGCGGCTTTGCCCGCGCGCTCACTGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCG 120  
QY 121 GGCACCTCCATCACTAGGGGTAAATCGGAAGCGCCTCCACCGCTGCGCGCTGACGCGTGA 180  
Db 121 GGCACCTCCATCACTAGGGGT-----TCCGTGAGGGGTGAGTCTGTA 163  
QY 181 CGTAAATTAAGTATAGG---GAGTGGTCTGTATTAGCTGTCACTGAGTGTCTTTGCG 237  
Db 164 CGTGAATTAAGTATAGGAGGTCTGTATTAGAGTCACTGAGTG-TTTTGC 222  
QY 238 GACATTTTGGCAGACCAACGCTGGCAATTTAGGGTATATATGCGCGAGTGAGCGAGCAT 297  
Db 223 GACATTTTGGCAGACCAATGTGTCACTGCGGTATTAAAGCCCGAGTGAGCAGCAGCGGT 282  
QY 298 CTCCATTTTG-ACCGGAAATTTGAACGAGCAGCAGCAGCATGCGGGCTTCTACGAGATCG 356  
Db 283 CTCCATTTTGAAGCGGGAGGTTTGAACGCGCAGCGCCATGCGGGGTTTACGAGATTG 342  
QY 357 TGATCAAGGTGCGGAGCGACCTGGAAGCAGCAGCAGCATTTCTGACTCGTTTGTGA 416  
Db 343 TGAATTAAGGTCCCGAGCAGCTTGACGGCATCTGCCCGCATTTCTGACAGCTTTGTGA 402  
QY 417 GCTGGGTGCGCGAGGAAGTGGAGCTGCCCCCGGATTCTGACATGATCTGAATCTGA 476  
Db 403 ACTGGGTGCGCGAGGAAGTGGAGTTCGCCGAGATTCTGACATGATCTGAATCTGA 462  
QY 477 TTGAAGAGGCAACCCCTGACCGTGGCGGAGAGCTGACGCGGCACTTCTGTGCAATGGC 536  
Db 463 TTGAAGAGGCAACCCCTGACCGTGGCGGAGAGCTGACGCGGCACTTCTGACGGAATGGC 522  
QY 537 GCGCGGTGAAGTAAAGCCCCCGAGGCGCTCTTTGTTGTTCAATTGAGAAGGGCGAGTCT 596  
Db 523 GCGGTGTAAGTAAAGCCCCCGAGGCGCTTTTCTTTGTTCAATTGAGAAGGGAGAGAGCT 582  
QY 597 ACTTCCACTCCATATTCTGTGAGAGCAGCGGGGTCAATTCATGTGTGCGGCGCT 656  
Db 583 ACTTCCACTGACAGTGTCTGTGAAACCAACCGGGGTGAATTCATGTGTTTGGGACGTT 642  
QY 657 TCTGAGTCAAGTTAGGACAAGCTGTGTGAGACCATCTACCGCGGGATCGAGCCGACCC 716  
Db 643 TCTGAGTCAAGTTGCGGAAAACTGATTCAAGAAATTACCGCGGGATCGAGCCGACTT 702  
QY 717 TGCCCACTGGTTGCGGTGACCAAGACGCGTAAATGCGCGGAGGGGGAACAAGGTGG 776  
Db 703 TGCCAACTGGTTGCGGTGACCAAGACCAAGAAATGCGCGGAGGGGGAACAAGGTGG 762  
QY 777 TGAGAGGTGTACATGCCCACTACTCTGCCCCAAGACTGACCCGAGCTGCACTGGG 836  
Db 777 TGAGAGGTGTACATGCCCACTACTCTGCCCCAAGACTGACCCGAGCTGCACTGGG 836

Db 763 TGAGTAGTGCTACATCCCAATTACTTGCTCCCAAAACCCAGCCTGAGCTCCAGTGGG 822  
QY 837 CGTGACTTAACATGAGAGATATATAAGCGCCTGTTGAACCTGGCGGCAACCGGC 896  
Db 823 CGTGACTAATATGAAACAGATTTAAGCGCCTGTTGAATCTACGAGCGTAACGGT 882  
QY 897 TCGTGCGGACAGCACCCTGACCCACGTCAGCCAGACCCAGAGCAGAACAGAGAAATCTGA 956  
Db 883 TGTGCGCGGACAGCATCTGACGACGTCGTGCGAGACCGGAGCAGAACAGAGAAATCAGA 942  
QY 957 ACCCCAATTCTGACGCGCGCTGTATCCGGTCAAAAACCTCCGCGCTACATGAGCTGG 1016  
Db 943 ATCCCAATTCTGATGCGCGGTGATCAGATCAAAAACCTTCAGCCAGGTACATGAGCTGG 1002  
QY 1017 TCGGTGCTGTGTGACCGGGGATCACTCCGAGAAAGTGTATCCAGAGAGACAGG 1076  
Db 1003 TCGGTGCTGTGTGACCAAGGGGATTACTCCGAGAAAGTGTATCCAGAGAGACAGG 1062  
QY 1077 CTTGTACATCTCTTCAAGCGCGCTTCCAACTGCGGCTCCAGATCAAGCGCTTGG 1136  
Db 1063 CTTATACATCTCTTCAATGCGGCTCCAACTGCGGCTCCAAATCAAGCTGCTTGG 1122  
QY 1137 ACAATGCCGGAAGATCATGGGCTGACCAAAATCCGCGCCGACTACTGTAGGCCCG 1196  
Db 1123 ACATGCGGGAAGATTAATGACCTGACTAAACCGCCCGACTACTGTGGGCCAGC 1182  
QY 1197 CTGCGCGCGGACATTAAACCAACCGCATCTACCGCATCTGAGCTGAACGGCTACG 1256  
Db 1183 AGCCCGTGAGGACATTTCCAGCAATCGGATTATATAAATTTTGAACATAACGGGTACG 1242  
QY 1257 AACCTGCTACGCGCGCTCGCTTCTCTGCGTGGGCCAGAAAAGTTGCGGAAGCGCA 1316  
Db 1243 ATCCCAATATGCGGCTTCGCTTCTCTGAGTGGGCCAGAAAAGTTGCGGAAGGGA 1302  
QY 1317 ACAACATCTGCTGTTTGGCGCGGCCAACACGCGGCAAGCAACATCGCGGAAGCCATCG 1376  
Db 1303 ACAACATCTGCTGTTTGGCGCTGCAACTACCGGGAAGCAACATCGCGGAGCCATAG 1362  
QY 1377 CCCAGCGGTGCGCTTCTACGGCTGCGTCACTGAGCAATGAGAACTTCCCTCAATG 1436  
Db 1363 CCCACACTGTGCGCTTCTACGGGTGCGTAACTGAGCAATGAGAACTTCCCTCAACG 1422  
QY 1437 ATTGCTCGACAAAGATGTGATCTGTGTGGAGAGGGCAAGATGACGGCCAGAGTCTGG 1496  
Db 1423 ACTGTGTGACAAAGATGTGATCTGTGTGGAGAGGGGAAGATGACGGCCAGAGTCTGG 1482  
QY 1497 AGTCCGCCAAGCCATTCTGCGCGGCGAGCAAGGTGCGGTGGAACCAAAAGTCAAGTCT 1556  
Db 1483 AGTCCGCCAAGCCATTCTGGAAGAGCAAGGTGCGGTGGAACCAAAATGCAAGTCT 1542  
QY 1557 CCGCCAGATGACCCCAACCCCGTGTGATGTCACCTCCAAACCAACATGTGCGCGCTGA 1616  
Db 1543 CCGCCAGATGACCCCAACCCCGTGTGATGTCACCTCCAAACCAACATGTGCGCGCTGA 1602  
QY 1617 TTGAAGGGAACAGCACCACTTTCAGACACGAGCGGTTGACGAGCGGATGTTCAAT 1676  
Db 1603 TTGAAGGGAACCTCAACGACCTTCGAACACGAGCGGTTGACGAGCGGATGTTCAAT 1662  
QY 1677 TTGAATCAACCCCGCTGTGAGCATGACTTTGGAAGGTGAACAAGCAGAGTCAAG 1736  
Db 1663 TTGAATCAACCCCGCTGTGATCATGACTTTGGAAGGTGAACAAGCAGAGTCAAG 1722  
QY 1737 AGTTCTTCGCTGGGCGGAGATCACTGACCGAGGTGCGCATGAGTTCTACGTCAAG 1796  
Db 1723 ACTTTTTCGGTGGGCAAGGATCACTGTTGAGGTGAGCATGAAATTTCTACGTCAAA 1782  
QY 1797 AGGCTGAGCCAAACAAGACCCCGCCCGATGACGCGGATTAAGCGAGCCCAAGCGGG 1856  
Db 1783 AGGCTGAGCCAAAGAAAGACCCCGCCCGATGACGCGGATTAAGTGAAGCCCAAGCGGG 1842  
QY 1857 CTTGCGCTCAGTGGGATCATGACGTGAGACGCGGAAGGAGCTCCGTTGACTTTG 1916  
Db 1843 TGGCGAGTCAAGTTGCGGAGCATGACGTGAGACGCGGA---AGCTTGATCACTACG 1899





Db 4045 CATCTGGGCAAGATTCACACACGAGGACATTTTACCCCTCTCCCTCATGGGTGG 4104  
Qy 4133 CTTTGACTCAGAACCCGCTCTCAGATTCCTCATCAAAAACAGCCTGTCTCCGAA 4192  
Db 4105 ATTGGACTTAAACACCTCTCTCCACAGATTCATCAAGAACCCCGGTACTCGGAA 4164  
Qy 4193 TCCTCCGCGGAGTTTTCAGCTACAAAGTTTCTTATTCATCACCAATATCTCCACAG 4252  
Db 4165 TCCTTCGACCACTTCAGTGCAGAAAGTTTCTTCTTATTCACAGTACTCCACGGG 4224  
Qy 4253 ACA-AGTGAAGTGAATGAATGAGAGTGCAGAAAGAAACAGCAAGCGTGAATC 4311  
Db 4225 ACACGCTCAGCTGAGATCGAGTGGAGCTGCAGAGAAAGAAACAGCAACGCTGAATC 4284  
Qy 4312 CCGAAGTCAGTACATCCAAATTAAGCAAAATCTGCCAAGCTGAATTTACTGTGACA 4371  
Db 4285 CCGAATTCAGTACACTTCCAACTACAAAGTCTGTTAATCGTGACTTAACCGTGATA 4344  
Qy 4372 ACAATGACTTTTACTGAGCCTCGCCCACTGGCACCCGTTACCTTACCCGCTCCCTGT 4431  
Db 4345 CTAATGGCGGTATTCAGAGCCTCGCCCACTGGCACAGATACCTGATCTGTAATCTGT 4404  
Qy 4432 AATTACGTGTTAATCAATAAACCGGTTGATTCGTTCAAGTTGAATTTGGTCTCTCTCC 4491  
Db 4405 AATTGCTTGTAAATCAATAAACCGTTTAAATTCGTTCAAGTTGAATTTGGTCTCTCTCC 4464  
Qy 4492 TTCTTATCTTATCGGTTTACCAGTGTATAGCTTACATTAACCTGCTGGTGGCTTCG 4551  
Db 4465 TTCTTCTTATCTAGTTTCCATGCTACGTAGATAGTAGCATGGCGGGTTAATCATTA 4524  
Qy 4552 CGATAAAGACTTACGTATCGGTTACCCCTAGTATGAGTGGCCCACTCCCTCTCTG 4611  
Db 4525 ACTACAAAGA-----ACCCTAGTATGAGTGGCCCACTCCCTCTCTG 4568  
Qy 4612 CGCGCTCGCTCGCTCGGTGGGGCTGCGGACCAAGGTCCGACAGCGGAGAGTCTGTCT 4671  
Db 4569 CGCGCTCGCTCGCTCACTAGAGCCGGGCGCAAAAGGTCCGCCGAGCGCCGGCTTGGCC 4628  
Qy 4672 CTGCGGCGCCCAACGAGCGAGCGGCGGAGAGGAGTGGGCAA 4718  
Db 4629 CGGGCGGCTCAGTAGCAGCGAGCGCGGAGAGAGGAGTGGGCAA 4675

RESULT 13  
AX205073  
LOCUS AX205073 7557 bp DNA linear PAT 30-AUG-2001  
DEFINITION Sequence 3 from Patent WO0155361.  
ACCESSION AX205073  
VERSION AX205073.1 GI:15394314  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 7557)  
AUTHORS Hardy, S.F.  
TITLE Recombinant aav packaging systems  
JOURNAL Patent: WO 0155361-A 3 02-AUG-2001;  
Chiron Corporation (US)  
FEATURES  
source 1.7557  
/organism="unidentified"  
/db\_xref="taxon:32644"  
/note="recombinant DNA"  
BASE COUNT 1970 a 1940 c 1953 g 1694 t  
ORIGIN

Query Match 63.4%; Score 2993.4; DB 6; Length 7557;  
Best Local Similarity 80.4%; Pred. No. 0;  
Matches 3633; Conservative 0; Mismatches 841; Indels 45; Gaps 9;

Qy 18 GCGCGCTCGCTCGCTCGGTGGGGCTCGGACCAAGGTCCGACAGCGGAGAGTCTGCT 77  
|||||

Db 4 GCGCGCTCGCTCGCTCACTGAGCGCCCGCGCAAGCCCGGCGTGGCGGACCTTGG 63  
Qy 78 TCTGCGGCCCCCAGCGAGCGAGCGCGCAGAGAGGAGTGGGCAACTCCATCACTAG 137  
Db 64 TCGCCCGGCTCAGTGAAGCGAGCGCGCAGAGAGGAGTGGGCAACTCCATCACTGA 123  
Qy 138 GGGTAAT-----CGGAAGCGCTCCACGCTGCCGCTC 172  
Db 124 TGGGACTTGTAGCATATCTGTATATGTATGTATATAGCAAGTATCCGAGGGGTGG 183  
Qy 173 AGCGCTGACGTAATTAAGTCAATAGG--GAGTGTCTGTATTAAGTGTACCTGAGT 229  
Db 184 AGTGTGACGTAATTAAGTCAATAGGTTAGGAGGTCTGTATTAAGGTCACTGAGT 243  
Qy 230 GCTTTGCGACATTTTGGCAGACCAAGTGGCATTAGGTTATATAGCCGAGTGAAGC 289  
Db 244 G-TTTTGGACATTTTGGCAGACCAATGTGTACGCTGGGTATTAAGCCGAGTGAGCA 302  
Qy 290 AGCAGATCTCCATTTG-ACCGGAAATTTGAACGAGACAGCAGCAGTCCGGCTTCTA 348  
Db 303 CGCAGGCTCTCCATTTTGAAGCGGAGGTTTGAACGCGCAGCGCCATGCGGGTTTGA 362  
Qy 349 CGAGATCGTATCAAGTGGCGGACGACCTGAGCAGACCTGCGGCAATTTCTGACTC 408  
Db 363 CGAGATGTGATTAAGTCCCGCAGCAGCTTGAACGAGCATCTGCCGCAATTTCTGAC 422  
Qy 409 GTTGTGACGTGGTGGCGGAGAGGATGGAGCTGCCCGGATTTCTGACATGATCT 468  
Db 423 CTTTGTGAACCTGGTGGCGGAGAGGATGGAGTGGCGGCAATTTCTGACATGATCT 482  
Qy 469 GAATCTGATTTGAGCAGGACCCCTGACCGTGGCCGAGAGCTGACCGGCACTTCTGT 528  
Db 483 GAATCTGATTTGAGCAGGACCCCTGACCGTGGCCGAGAGCTGACCGGCACTTCTGTAC 542  
Qy 529 CCAATGCGCGCGCGTGAAGGCGCGGAGGCGCTTCTTGTGATTCAGTTGAGGAGG 588  
Db 543 GGAATGCGCGCGTGAAGGCGCGGAGGCGCTTCTTGTGCAATTTGAGGAGG 602  
Qy 589 CGAATCTACTTCCACCTCAATATCTGTGAGAGACCAAGCGGCTCAATCCATGCTGT 648  
Db 603 AGAGAGCTACTTCCACATGACGTGCTGTGAAACCAAGCGGCTGAATCCATGCTT 662  
Qy 649 GGGCGCTTCTGAGTCAAGTATAGGAGCAAGCTGTGACAGCATCTACCGCGGATCGA 708  
Db 663 GGGAGCTTCTGAGTCAAGTATAGGAGCAAGCTGTGACAGCATCTACCGCGGATCGA 722  
Qy 709 GCGGACCTTCCCAACTGTTGCGGCTGACCAAGACCGGTAATGGCGCGGAGGAGGAA 768  
Db 723 GCGGCTTTCGCAAACTGTTGCGGCTGACCAAGACCAAGATGGCGCGGAGGAGGAA 782  
Qy 769 CAAGTGTGAGCAGTGTCAATCCCACTACTCTGCGGCAAGTCAAGCCGAGCT 828  
Db 783 CAAGTGTGAGTGTGATGATCAATCCCACTACTCTGCGGCAAGTCAAGCCGAGCT 842  
Qy 829 GCAGTGGCGGTGAGTAACTGAGAGGATATATTAAGCGCTGTTGAAGTGGCGGAGCG 888  
Db 843 CCAGTGGCGGTGAGTAACTGAGAGGATATATTAAGCGCTGTTGAAGTGGCGGAGCG 902  
Qy 889 CAAAGCGCTGCTGGCGGACGACTGACCCAGCTGACGACAGCCAGAGAGCAACAAGA 948  
Db 903 TAAACGTTGTTGGCGGACGACTGACGACGCTGTCCGACGACGAGAGCAACAAGA 962  
Qy 949 GAATCTGAACCCCAATTTGACGCGCTGTCAATCCGCTCAAAAACCTCGCGGCTACAT 1008  
Db 963 GAATCAGAAATCCCAATTTGATGCGCGGTGATCAGATCAAAAACCTTACGCGGCTACAT 1022  
Qy 1009 GGAGCTGTGGGTGGGTGGGTGAGCGGGGATCACTCCGAGAGAGAGTGTATCCAGGA 1068  
Db 1023 GGAGCTGTGGGTGGGTGGGTGAGCAAGGGGATTACTCGAGAGAGAGTGTATCCAGGA 1082  
Qy 1069 GGACGAGGCTCTGATCATCTCTTCAAGCGCGCTTCAACTGCGGCTCCAGATCAAGGC 1128  
Db 1083 GGACGAGGCTCTATACATCTCTTCAATGCGGCTCCAACTCGCGGCTCCAAATCAAGGC 1142



QY 1129 CGCTTGACATGCGGCAAGATCATGGCGGTGACCAATCCGCGCCGACTACCTGT 1188  
DB 1143 TGCTTGACATGCGGAAAGATTATGAGCTGACTAAACCGCCCGACTACCTGT 1202  
QY 1189 AGGCCCCGCTCCGCGCGGACATTAAACCAACCGCATCTACCGCATCTGAGCTGAA 1248  
DB 1203 GGGCCAGCAGCCCGTGAAGACATTTCAGCAATCGAATTATTAATTTGAACTAAA 1262  
QY 1249 CGGCTACGAACCTGCTTACGCGGCTCCGCTTTCTCGGCTGAGCCCAAGAAAGTTCCG 1308  
DB 1263 CGGTACGATCCCAATATGCGGCTTCGCTTTCTGGGATGAGCCAGAAAGTTCCG 1322  
QY 1309 GAAGCGCAACCACTCTGGCTGTTTGAGCGGCAACGCGCAAGACCAACATCGCGGA 1368  
DB 1323 CAAGAGAACACCACTCTGGCTGTTTGAGCGGCAACGCGCAAGACCAACATCGCGGA 1382  
QY 1369 AGCCATCGCCCAACGCGGCTCCCTTCTACGGCTGCTCACTGACCAATGAACTTTCC 1428  
DB 1383 GGCCTAGCCCACTGCTGCTCCCTTCTACGGGTCGTAACTGACCAATGAACTTTCC 1442  
QY 1429 CTTCATGATGTCGTCGACAAAGATGATGTCGTGGGAGGAGGCAAGATGACGGCCA 1488  
DB 1443 CTTCACGACTGTGTGACAAAGATGATGTCGTGGGAGGAGGCAAGATGACGGCCA 1502  
QY 1489 GGTCTGAGTCCGCAAGGCCATTCTCGCGGCAAGAGTGCGCGTGAACCAAAAGTG 1548  
DB 1503 GGTCTGAGTCCGCAAGGCCATTCTCGAGGAAGCAAGTGCGCGTGAACCAAAAGTG 1562  
QY 1549 CAAGTCGTCGCGCAGATCGACCCCAACCCCGTGAATGTCACCTCCAACCAACATGTG 1608  
DB 1563 CAAGTCGTCGCGCAGATCGACCCCAACCCCGTGAATGTCACCTCCAACCAACATGTG 1622  
QY 1609 CGCCGTGATGACGGGAACAGCACCACTTCGAGCACAGCAGCCGTTGCAAGCCGAT 1668  
DB 1623 CGCCGTGATGACGGGAACAGCACCACTTCGAGCACAGCAGCCGTTGCAAGCCGAT 1682  
QY 1669 GTTCAATTTGAATCAACCCCGCTCTGAGCATGACTTTGCAAGGTGACAAAGCAGA 1728  
DB 1683 GTTCAATTTGAATCAACCCCGCTCTGAGCATGACTTTGCAAGGTGACAAAGCAGA 1742  
QY 1729 AGTCAAGAGTCTTCGCTGGCGCGCAGGATCACTGACCGAGTGCGCATGAGTTCTA 1788  
DB 1743 AGTCAAGAGTCTTCGCTGGCGCGCAGGATCACTGAGTTGAGGTGAGCATGAATTTCTA 1802  
QY 1789 CGTCAGAAAGGTGAGCAACAAAGACCCCGCCGATGACGCGGATAAAGCGAGCC 1848  
DB 1803 CGTCAGAAAGGTGAGCAACAAAGACCCCGCCGATGACGCGGATAAAGCGAGCC 1862  
QY 1849 CAAGCGGCGCTCCCTCAGTCGCGGATCCATCGACGTGACGCGGAAGAGCTCCGT 1908  
DB 1863 CAAGCGGCGCTCCCTCAGTCGCGGATCCATCGACGTGACGCGGAAGAGCTCCGT 1919  
QY 1909 GGAATTTGCGCAGGTACCAAAACAATGTTCTGTCACGCGGATGCTTCAGATGCT 1968  
DB 1920 CAATACGACAGACAGTACCAAAACAATGTTCTGTCACGCGGATGCTTCAGATGCT 1979  
QY 1969 GTTTCCTGCAAGACATGCGAGAGATGATCAATTTCAACATTTGCTTCACGACGG 2028  
DB 1980 GTTTCCTGCAAGACATGCGAGAGATGATCAATTTCAACATTTGCTTCACGACGG 2039  
QY 2029 GACGAGAGACTGTTCAAGTGTCTCCCGCGGTGTCAGAAATCTCAACCGGT---CGTCAG 2085  
DB 2040 ACAGAAAGACTGTTAAGTGTCTTCC---CGTCAGAAATCTCAACCGGTCTGTGCT 2096  
QY 2086 AAAGAGACGTATCGAAACTCTGTGCATTCATCTCTGCGGCGCGGCTCCGAGAT 2145  
DB 2097 CAAAAAGCGTATCAGAACTGTGTACATTCATCATATGAGGAAAGGTGCGAGA--- 2153  
QY 2146 TGCTTGCTCGGCTGCGATCTGTCAACGTGACCTGATGACTGTGTTCTGAGCAATA 2205  
DB 2154 CGCTTGCACTGCTGCGATCTGTGATGATGATGATGATGATGATGATGATGATGATGAT 2213

QY 2206 AATGACTTAAACAGGTATGCTGCGGATGTTATCTTCAGATTTGGCTCGAGGACCAAC 2265  
DB 2214 AATGATTTAAATCAGTATGCTGCGGATGTTATCTTCAGATTTGGCTCGAGGACACTC 2273  
QY 2266 TCTTGAGGCAATTCGAGTGTGAGGACTTTGAAACCTGAGACCCCGAAAGCCCAAGCCA 2325  
DB 2274 TCTTGAGGCAATTCGAGTGTGAGGACTTTGAAACCTGAGACCCCGAAAGCCCG 2333  
QY 2326 ACCAGCAAAAGCAGACGACGCGCGGCTGTGTGCTTCTGCTTCAAGTACCTCGAC 2385  
DB 2334 CAGACGCGCAATAAGACGACGACGCGGCTGTGTGCTTCTGCTTCAAGTACCTCGAC 2393  
QY 2386 CTTTCAACGCACTCGACAAAGGAGCCCGTCAACGCGGCGGACGACGCGGCTCGAC 2445  
DB 2394 CTTTCAACGCACTCGACAAAGGAGCCCGTCAACGAGGACGACGCGGCTCGAC 2453  
QY 2446 ACGACAAAGCCTACGACGACGCTCAAGCGGCTGACATCCGTACCTGCGTATTAAC 2505  
DB 2454 ACGACAAAGCCTACGACGCGGCTGACGACGAGACCAACCGTACCTCAAGTACCAAC 2513  
QY 2506 ACGCGACCGGAGTTTCAAGAGCGCTCTGCAAGAAATACGTTTGGGCGCAACCTCG 2565  
DB 2514 ACGCGACCGGAGTTTCAAGAGCGCTTAAAGAAATACGTTTGGGCGCAACCTCG 2573  
QY 2566 GCGGAGCTCTTCCAGGACCAAGAGCGGCTTCTGAACTCTCGGCTGCTGTTGAGGAG 2625  
DB 2574 GACGAGAGCTTCTCCAGGCGAAAGAGGCTTCTGAACTCTGCGGCTGTTGAGGAG 2633  
QY 2626 GCGTAAAGCGGCTCTGAAAGAAAGCTCCGCTAGAGAGCTGCGCAAGAGCGAGACT 2685  
DB 2634 CTGTTAAGAGCGGCTCCGGAAGAAAGAGCGGCTAGAGAGCTCTCTGTGAGCGAGACT 2693  
QY 2686 CCTCTCGGCACTCGGCAAGACAGCGGCTTCAAGAGCGGCTTCAAGAGCGGCTTCAAG 2745  
DB 2694 CCTCTCGGCACTCGGCAAGAGCGGCTTCAAGAGCGGCTTCAAGAGCGGCTTCAAG 2753  
QY 2746 AGACTGCGACTCAGAGTCACTCCCGATCCACAACTCTCGGAGAACTCCAGCAACCC 2805  
DB 2754 AGACTGCGACTCAGAGTCACTCCCGATCCACAACTCTCGGAGAACTCCAGCAACCC 2813  
QY 2806 CCGCTGCTGAGGACTTACAAATGCTTCAAGCGGCTGCGGCAACCAATGCGAGCAATA 2865  
DB 2814 CCGCTGCTGAGGACTTACAAATGCTTCAAGCGGCTGCGGCAACCAATGCGAGCAATA 2873  
QY 2866 ACGAAGCGCGGAGCGAGTGGGTAATGCTCAGAAATTTGCAATGCGATTCACATGCG 2925  
DB 2874 ACGAAGCGCGGAGCGAGTGGGTAATGCTCAGAAATTTGCAATGCGATTCACATGCG 2933  
QY 2926 TGGGCGACAGTCAACCAACGACCGGCTGCGGCTGCGGCAACCAATGCGATTCAC 2985  
DB 2934 TGGGCGACAGTCAACCAACGACCGGCTGCGGCTGCGGCAACCAATGCGATTCAC 2993  
QY 2986 ACCTTCAACCAATCTCCAGTGTCTTCAACGCGGCGGCGGCAACCAACCACTACTTCG 3045  
DB 2994 ACCTTCAACCAATCTCCAGTGTCTTCAACGCGGCGGCGGCAACCAACCACTACTTCG 3050  
QY 3046 GCTACAGACCCCTGCGGATTTTGAATTTCAACAGATTTCCACTGCTTTTCAACCA 3105  
DB 3051 GCTACAGACCCCTGCGGATTTTGAATTTCAACAGATTTCCACTGCTTTTCAACCA 3110  
QY 3106 GTGACTGCGACGCTCATCAACCAATTTGGGATTTCCGCGCAAGAGACTCAACTTCA 3165  
DB 3111 GTGACTGCGACGCTCATCAACCAATTTGGGATTTCCGCGCAAGAGACTCAACTTCA 3170  
QY 3166 AACTCTTCAACATCCAGTCAAGAGTCAAGAGATGATGCGTCACAACCATGCTA 3225  
DB 3171 AACTCTTCAACATCCAGTCAAGAGTCAAGAGATGATGCGTCACAACCATGCTA 3230  
QY 3226 ATAACCTTACAGCAGCGTTTCAAGTCTTCTGCACTCGAGTACAGCTTCCGATGCTC 3285  
DB 3231 ATAACCTTACAGCAGCGTTTCAAGTCTTCTGCACTCGAGTACAGCTTCCGATGCTC 3290  
QY 3286 TCGGCTGCGGACCAAGGCTGCTTCCCTTCCGTTCCGCGGAGCGTGTATGATTCGCG 3345

Db 3291 TCGGCTCGGCGATCAAGATGCCCTCCCGCGCTTCCACAGACAGCTTCATGTCGCCAC 3350  
 Oy 3346 AATACGGCTACTGACGCTCAACAATGGACCAAGCCGTGGACCTTCATCTTTACT 3405  
 Db 3351 AGTATGATACCTCAACCTGACACACGGGAGTCAGGCAAGTAGGACGCTTCATTTACT 3410  
 Oy 3406 GCCTGGAATATTTCCCTTCTCAGATGCTGAGAACGGGCAACAACCTTACCTTCAGCTACA 3465  
 Db 3411 GCCTGAGTACTTCTTCTTCTCAGATGCTGAGTACCGGAACAACCTTACCTTCAGCTACA 3470  
 Oy 3466 CCTTTGAGGAAGTGCCTTTCCACAGACGCTACGGCCACAGAGCTTCGACCGGCTGA 3525  
 Db 3471 CTTTGTAGGACGTTCTTTCACAGACGCTACGCTCACAGCCAGAGTCTGACCGTCTCA 3530  
 Oy 3526 TGAATCCTCTCATCGACCAATACCTGTATTAACCTGAACAGAACTCAAAATCAGTCCGAA 3585  
 Db 3531 TGAATCCTCTCATCGACCAATACCTGTATTAACCTGAACAGAACTCAAAATCAGTCCGAA 3590  
 Oy 3586 GTGCCCAAAACAAGACTTGTCTTACGCGTGGGTCTCCAGCTGGCATGTCTGTACG 3645  
 Db 3591 CCACCACGCGATCAAGGCTTCAAGTCTTCTCAGGCGGAGCGAGTGCATTCGGGACCAGT 3650  
 Oy 3646 CCAAAAACCTGCTACCTGACCCCTGTATTCGGCAGCAGCGGCTTCTAAACAAAAACAG 3705  
 Db 3651 CTAGGAACCTGCTTCTCTGACCCCTGTATTCGCGCAGCAGGAGTATCAAGACATCTGCGG 3710  
 Oy 3706 ACAACAACAACAGCAATTTTACCTGAGCTGTCTTCAAAATATTAACCTCAATGGCGCTG 3765  
 Db 3711 ATAAACAACAAGTAACTCTGAGCTGAGCTTACCAAGTACCACTCAATGGCAGAG 3770  
 Oy 3766 AATCCATCATCAACCTTGGCACTGTATGCGCTTACACAAGACGACGAAGCAAGTCTT 3825  
 Db 3771 ACTCTGTGTAATCCGGGCGCGGCAATGCAAGCCACAAGACGATGAAGAAAGTTT 3830  
 Oy 3826 TTCCCATGAGCGGTGTATGATTTTGAAGAAAGAGCGCGGAGCTTCAACACTGCAT 3885  
 Db 3831 TTCTCTAGAGCGGGGTCTCATCTTTGGGAGCAAGGCTCAGAGAAAAACAATGTGACA 3890  
 Oy 3886 TGGACAATGTCATGATTACAGACGAAGAGAAATTAAAGCCATTAACCTGTGGCCACCG 3945  
 Db 3891 TTGAAGAGTCAATGATTACAGACGAAGAGAAATCAGACAAACCAATCCCGTGTACCGG 3950  
 Oy 3946 AAAGATTGGGACCGTGGCAGTCAATTTCCAGACGACGACGACGACCCCTGCGACCGGAG 4005  
 Db 3951 AGCAGTATGCTTCTGTATCTACCAACTTCCAGAGAGGCAACAGACAGACTACCGCAG 4010  
 Oy 4006 ATGTGATGCTATGGAGCATTACCTGGCATGTGTGGCAAGATAGAGACGTGTACCTGC 4065  
 Db 4011 ATGTCAACACAAAGCGCTTCTTCCAGGATGTGTGGCAGACAGAGATGTGTACCTTC 4070  
 Oy 4066 AGGGTCCCATTTGGGCAAAATTCCTCAACAGATGACACTTTCAACCCCTCTCTCTTA 4125  
 Db 4071 AGGGGCCATCTGGGCAAGATTCCACACGAGACGACATTTTCAACCCCTCTCTCTCA 4130  
 Oy 4126 TGGGCGGCTTTGGACTCAAGAACCCGCTCTCAGATCCTCATCAAAAACAGCGCTGTC 4185  
 Db 4131 TGGGTGATTCGGAATTAAACACCCCTCTCTCAACAGATTCTCATCAAGAACACCCCGGTAC 4190  
 Oy 4186 CTGCGAATCTCCGGCGGAGTTTCAAGTCAAAAGTTGCTTCAATTCATCAACCAATACT 4245  
 Db 4191 CTGCGAATCTTCCAGCACCTTCAAGTGGGCAAAAGTTGCTTCTTCAATCAACAGTACT 4250  
 Oy 4246 CCACAGCAAGTGAAGTGTGAATTTGAATGGAGCTGACAGAAAGAAACAGCAAGCGCT 4305  
 Db 4251 CCACGGGACAGTCAAGCTGAGATGAGTGGAGCTGACAGAGGAAACAGCAAGCGCT 4310  
 Oy 4306 GGAATCCGGAAGTGAAGTCAACATCAATTAATGCAAAATCTGCCAAGCTGATTTTACTG 4365  
 Db 4311 GGAATCCGGAATTCAGTACACTTCAACTACAACAAGTCTGTTAATGTGACTTTACTG 4370  
 Oy 4366 TGGACAACATGACTTTATCTGAGCCTCGCCCATTTGACCCGTTTACCTTACCCGTC 4425

Db 4371 TGGACAATATGCGGTGTATTCAGAGCCCTGCCCCATTTGGCACCAGATACCTGACTGTA 4430  
 Oy 4426 CCCTGTAATTAAGTGTGTAATCAATAAACCGGTGATTCGTTCACTTGAACCTTGTCTC 4485  
 Db 4431 ATCTGTAATGCTTGTGTAATCAATAAACCGTTAATTCGTTTCACTTGAACCTTGTCTC 4490  
 Oy 4486 CTGCTCTTATCTTATC 4504  
 Db 4491 TGGGTATTTCTTCTTATC 4509

RESULT 14  
 AF369963  
 LOCUS AF369963 7327 bp DNA circular SYN 22-JUL-2001  
 DEFINITION Cloning vector pAAV-RC, complete sequence.  
 ACCESSION AF369963  
 VERSION AF369963.1 GI:14994089  
 KEYWORDS  
 SOURCE Cloning vector pAAV-RC.  
 ORGANISM Cloning vector pAAV-RC  
 REFERENCES 1 (bases 1 to 7327)  
 AUTHORS Grafsky, A.J. III.  
 JOURNAL PAAV-RC for AAV Helper-Free System  
 TITLE Unpublished  
 REFERENCE 2 (bases 1 to 7327)  
 AUTHORS Grafsky, A.J. III.  
 TITLE Direct Submission  
 JOURNAL Submitted (16-APR-2001) Technical Services, Stratagene, 11011 N. Torrey Pines Rd., La Jolla, CA 92037, USA

FEATURES  
 source  
 1. 7327  
 /organism="Cloning vector pAAV-RC"  
 /db\_xref="taxon:161364"  
 /lab\_host="Escherichia coli"  
 /note="used for recombinant DNA processing"  
 131. 1996  
 /note="AAV-2 rep"  
 /codon\_start=1  
 /transl\_table=11  
 /product="adeno-associated virus 2 rep protein"  
 /protein\_id="AAK76418.1"  
 /db\_xref="GI:14994090"  
 /translation="MPGFYEIVIKVPSLDLHLEGISDSFVNWVAEKEMWELPDSMD  
 LNLIFQAPLTVAEKLQDFLEWTRVSKAPEALFVQFEKESYFHMVLEVTGKVS  
 MVLGRFLSQIREKLQRIYRGIEPTLPWFATVTRNAGGKNVDECIYPIVLLPK  
 TQPELQWAMTMEQYLISACNLTERKRLVAQHLTHVSQTOENKENQNPNSAPVIRS  
 KTSARYMELVGMVVDKGTSEKOWIOEDQASYISFNAASRSQIKALDNGKIMSL  
 TKTAPDYLVGQDPVEDISSNRIYKILELNGYDPOYASVFLGATKKFGKRNITWLF  
 PATGKTNIABEIAHTVPEFYGCNVNTNENFPENDCVDMKVIWEEGKMTAKVESAKA  
 ILGSKVRVDQKSSAQIDPTPIVITSNTNCAVIDNSTTFEHOQPLQDMFKFEL  
 TRRLDHDGKVTYKQVKDFFRMADKHVEVEHEFYVKKGAKRPAIPSDADISEPKRV  
 RESVAQPSSTDAEASINYADRYQNKCSRHVGMMMLFPCROCEMNONSNTCTHGO  
 DLECFPVSESQPVSVKAYQKLCYIHHIMGVKVPDACTACDLVNVLDLDCIFEO"  
 2013. 4220  
 /note="cap VP1"  
 /codon\_start=1  
 /transl\_table=11  
 /product="capsid protein VP1"  
 /protein\_id="AAK76419.1"  
 /db\_xref="GI:14994091"  
 /translation="MAADGYLPDWLEDTLSEGIROWMKLKPGRPPPKPAERHSDSRG  
 LVLPGYKYLGFENGLDKGEFVNEDAAALAEHKAAYDRQDSDGNPYLKYNHADA  
 RLKEDTSFGNLEGRAVFOAKRVLPLGLVEEPVKTAPGKRPVEHSFVEPSSSGTG  
 KAGQOPARKRLNFGOTGDADSVDPQPLGPPAAPSGLGTNTMATGSGAPADNNEGA  
 DGVNSSGNHCHDSITWMDPTSTRWALPTYNHLYKOISSOSGASNDNHYFGYS  
 TPWGYFDENRCHHSRPDMORLINNNGFRPKRLNFKLFIQVKEVTQNDGTTIAN  
 NLSTVOVFTDSEYQPYVLGSAHOGCLPPPADVEMVPOYGLTLNNGSOAVGRSSF  
 YCLEYFSPQMLRTGNFTFSYTFEDVPFHSYAHQSLSLRLMPLIDQYLYLSRTNT  
 PSGETTQSLRLOFSQAGASDIRQSRNWLPGPCYRQQRKSVKTSADNNNSYMTGATKY  
 HUNGRDSLVPNPGPAMASHKDEEKFEPQSGVLI FGQSGSEKTNVDIEKWI TDEEIR  
 TTNPVATEQYGSVETNLQGRNQAATPVNTQGVLPGMVQWDRDVILOGIWAKIPHT  
 DGHFHPSPLMGFGFKHPPIQLIKNTIPVPANPSTTFSAAKFASTIQTSGOVSEI

```
rep_origin      EWELOKENSKRWNPEIQYTSNYSNVNVDFTVDINGVSEBPIGTRYLTRNL"
/notes=".5161
/notes="f1 origin"
5292..6152
/notes="confers ampicillin resistance"
/codon_start=1
/cranel_table=11
/product="beta-lactamase"
/protein_id="AAK76420.1"
/db_xref="GI:14994092"
/cranelation="MSIQHFRVALIPEFAAFLPVFAHPETLVKVKDAEDQLGARVGY
IQLDLSGKILSFPRPEERFPMGTFKVLICGAVLSRIDAGEQLGRIRHYSQNDLVE
YSPVTEKHLTDGMTVRELCSAATMSDNTAANLLLTIGSPRELTAFLHMGDHTRL
DRWEPELNEAIPNDERDTMPVAMATLRLTGLTGLTASQQLIDMEADKVAQPL
LRSLAPAGWFIADKSGAGERGSRGIIAALGPDGKPSRIIVITTTGSQATMDENRQIA
EIGASLIKHM"
rep_origin      6161..7019
/notes="Colel origin"
BASE COUNT      1903 a      1887 c      1912 g      1625 t
ORIGIN
Query Match      61.0%; Score 2878.4; DB 12; Length 7327;
Best Local Similarity 81.3%; Pred. No. 0;
Matches 3403; Conservative 0; Mismatches 766; Indels 15; Gaps 5;
QY      324 GAGCAGCAGCCATGCGGGGCTTCTACAGATCGTGAATCAAGTGCCGAGCAGCTTGACG 383
Db      120 GCGCAGCGCCCATGCGGGGTTTACGAGATTGATTAAGTCCCGCAGCAGCTTGACG 179
QY      384 AGCACCCTGCGGGGATTTCTGACTCGTTTGATGCTGGGTGGCCGAGAGAAATGGAGC 443
Db      180 AGCATCTGCCCGGCAATTCTGACAGCTTGTGAAGTGGGTGGCCGAGAGAAATGGAGT 239
QY      444 TGCCCCCGGATTTGACATGATCTGAATCTGAATGAGCAGGCAACCCCTGACCCGTGGCCG 503
Db      240 TGCCGCGAGATTGACATGATCTGAATCTGAATGAGCAGGCAACCCCTGACCCGTGGCCG 299
QY      504 AGAAGCTGACGCGGCACTTCTGTGTCGAATGCGCGCGGTGAGTAAGGCCCGGAGGCC 563
Db      300 AGAAGCTGACGCGGCACTTCTGACGGAATGCGCGCGGTGAGTAAGGCCCGGAGGCC 359
QY      564 TCTTCTTTGTTCACTTCGAGAAGGGCGAGTCTTCCACTTCCACTTCTGCTGAGAGA 623
Db      360 TTTTCTTTGTTCACTTCGAGAAGGGCGAGTCTTCCACTTCCACTTCTGCTGAGAGA 419
QY      624 CCACGGGGGTCAATTCATGCTGCTGGGGCGCTTCTGAGTCAAGATTAGGACAAGCTGG 683
Db      420 CCACGGGGGTGAATTCATGCTGCTGGGGCGCTTCTGAGTCAAGATTAGGAAAACTGA 479
QY      684 TGCAGACCATCTACGCGGGGATCGAGCCGACCTGCCCACTGTTCCGGGTGACCAAGA 743
Db      480 TTCAGAGAATTACGCGGGGATCGAGCCGACCTGCCCACTGTTCCGGGTGACCAAGA 539
QY      744 CGCGTAATGCGCGCGGAGGGGGAACAAGTGTGTGACGAGTGTACTACATCCCACTACC 803
Db      540 CCAGAAATGCGCGCGGAGGGGGAACAAGTGTGTGATGAGTGTACTACATCCCACTACT 599
QY      804 TCCTGCCAAGACTCAAGCCGAGCTGAGTGGCGTGAATACATGAGAGATATATAA 863
Db      600 TGCTCCCAAAACCAGCTGAGCTCCAGTGGCGGTGAGTATATATGAGACATATTAA 659
QY      864 GCGCCTGTTGAACCTGCGCGGAGCGCAACCGGCTGTGGCGCAGCACTTGACCCACGTCA 923
Db      660 GCGCCTGTTGAATCTCAAGGAGCGTAAACGTTGTGTGGCGCAGCATCTGACGACAGTGT 719
QY      924 GCCAGACCCAGAGAGAGAAACAAGAGAACTGAACCCCAATCTGACGCGCTGTCAATCC 983
Db      720 CGCAGACGAGAGAGAGAAACAAGAGAACTGAATCCCAATCTGATGCGCGGTGATCA 779
QY      984 GGTCAAAAAACCTCGCGCGCTACATGAGCTGTGGGTGGCTGTGTGACCGGGGATCA 1043
Db      780 GATCAAAAAACCTCAAGCAGGTACATGAGCTGTGGGTGGCTGTGTGACCAAGGGATTA 839
```

```
QY      1044 CTTCCGAGAAAGCAGTGGATCCAGAGAGACCAAGCCCTGTTACATCTCTTCAACGCCGCTT 1103
Db      840 CTTCCGAGAAAGCAGTGGATCCAGAGAGACCAAGCCCTCATACATCTCTTCAATGCCGCTT 899
QY      1104 CCAACTCGCGGTCCAGATCAAGCCGCTCTGACAAATGCGGCAAGATCATGGCGCTGA 1163
Db      900 CCAACTCGCGGTCCCAATCAAGCTGCTTGGACAATGCGGAAAGATTATGAGCTGA 959
QY      1164 CCAAAATCCGCGCCGACTACTGTTAGGCCCCGCTCCGCGCGGACATTAACAACCAAC 1223
Db      960 CTAACAACGCGCCCGACTACTGTTGGCCAGCAGACCCGCTGAGAGACATTTCCAGCAATC 1019
QY      1224 GCATCTAACCGCATCTCTGAGCTGAACGGCTACGAACCTGCTACGCCGGCTCCGCTTTC 1283
Db      1020 GGATTTATAAATTTTGAACATAACGGGTACGATCCCAATATGCGGCTCCGCTTTC 1079
QY      1284 TCGGCTGGGCCCCAGAAAAGTTGGGGAAGCGCAACACCATCTGGCTGTTGGCGCGCA 1343
Db      1080 TGGGATGGGCCCAAGAAAAGTTGGGCAAGAGGAACACCATCTGGCTGTTGGCGCTGCA 1139
QY      1344 CCACGGGCAAGACCAACATCGCGGAAGCCATCGCCCCAGCGCTGCTTCAACGGCTGCG 1403
Db      1140 CTACCGGGAAGACCAACATCGCGGAGCCATAGCCACACTGTGCCCTTCAACGGGTGCG 1199
QY      1404 TCAACTGACCAATGAGAACTTCCCTTCAATGATTGCTGACAAAGATGATGTTGT 1463
Db      1200 TAAACTGACCAATGAGAACTTCCCTTCAACGACTGTGTCACAAGATGATGTTGT 1259
QY      1464 GGAAGAGGGCAAGATGACGCGCAAGTGTGAGTCCGCCAAGGCCATTTCTCGCGGCA 1523
Db      1260 GGAAGAGGGGAAGATGACCGCAAGTGTGAGTCCGCCAAGGCCATTTCTCGAGAGAA 1319
QY      1524 GCAAGTCCGCGGTGACCAAAAGTCAAGTGTCCGCCAAGTCGACCCACCCCGCTGA 1583
Db      1320 GCAAGTCCGCGGTGACCAAAAGTCAAGTGTCCGCCAAGTAGACCCGACTCCCGTGA 1379
QY      1584 TCGTCACTCCCAACACCAACATGTGCGCCGCTGATTGACGGGAACACACACTTGCAGC 1643
Db      1380 TCGTCACTCCCAACACCAACATGTGCGCCGCTGATTGACGGGAACCTCAACGACTTGCAGC 1439
QY      1644 ACCAGAGCGCTTGACAGACCGGATGTTCAATTTGAATCAACCCGCGCTGAGGATG 1703
Db      1440 ACCAGAGCGCTTGACAGACCGGATGTTCAATTTGAATCAACCCGCGCTGAGGATG 1499
QY      1704 ACTTGGCAAGTGAACAAGCAGAGTCAAAAGATTCTTCCGCTGGCGCAGGATCAG 1763
Db      1500 ACTTGGCAAGTGAACAAGCAGAGTCAAAAGATTCTTCCGCTGGCGCAGGATCAG 1559
QY      1764 TGACCGAGGTGGCGCATGAGTTCTACGTCAAAAGGTGAGCCAAACAAGACCCGCC 1823
Db      1560 TGAGTGAAGTGAAGCATGAAATCTCAAAAAGGTGAGCCAAAGAAAGACCCGCC 1619
QY      1824 CCGATGACGGGATTAAGAGCGACCGGCGCTGCCCTCAGTGGCGGATCCATCGA 1883
Db      1620 CCAAGTGAAGCATATTAAGTGAAGCCAAAGCGGTGCGGAGTGTGCGGACCATCGA 1679
QY      1884 CGTCAAGCGCGGAAGAGCTCCGGTGAATTTGCCGACAGGTACCAAAACAATGTTCTC 1943
Db      1680 CGTCAAGCGCGGAAGAGCTCCGGTGAATTTGCCGACAGGTACCAAAACAATGTTCTC 1736
QY      1944 GTACAGCGCGGATGCTTCAAGATGCTGTTCCCTGCAAGACATGCGAGAGAAATGAATCAGA 2003
Db      1737 GTACAGCGCGGATGATCTGATGCTGTTCCCTGCAAGACATGCGAGAGAAATGAATCAGA 1796
QY      2004 ATTCAACATTTGCTTCAAGCAGCGGACGAGAGCTGTTCAAGTGTCTCCCGCGGTGT 2063
Db      1797 ATTCAACATTTGCTTCAAGCAGCGGACGAGAGCTGTTAAGTGTCTCCCGCGGTGT 1853
QY      2064 CAGAAATCAACCGGT--CGTCAAAAGAGAGAGATGATGGAATCTGTGCAATTCATC 2120
Db      1854 CAGAAATCAACCGGTCTGTGTCAAAAAGCGTATCAGAAACTGTGCTACATTCATC 1913
QY      2121 ATCTGCTGGGCGCGGCTCCCGAAGATTGCTGTGCTGGCGCTGCAATCTGTCAACGTGACC 2180
```





Qy 4341 AAATCTGCCAAGCTGATTCTTACTGTGACAAATGACTTTATATCTGAGCCTCGCCCC 4400  
Db 4128 AAGTCTGTTAATGAGACTTACTGTGACACTAATGAGCGGTATTCAGAGCCTCGCCCC 4187  
Qy 4401 ATTGGACCCCGTTACCTTACCCCGCTGTAATTAACGTGTTAATCAATAAACCGGTTGA 4460  
Db 4188 ATTGGACACAGATACCTGACTCGTAATCTGTAATGCTTGTAAATCAATAAACCGTTTAA 4247  
Qy 4461 TTGCTTCAAGTTGAACTTGTGCTCTGCTCTGCTTCTTATCTTATC 4504  
Db 4248 TTGCTTCAAGTTGAACTTGTGCTCTGCTGATTTCTTCTTATC 4291  
RESULT 15  
AVU89790 4767 bp DNA circular VRL 21-AUG-1997  
LOCUS Adeno-associated virus 4, complete genome.  
ACCESSION U89790  
VERSION U89790.1 GI:2337938  
KEYWORDS  
SOURCE adeno-associated virus 4.  
ORGANISM adeno-associated virus 4.  
REFERENCE 1 (bases 1 to 4767)  
AUTHORS Chiorini, J.A., Yang, L., Liu, Y., Safer, B. and Kotin, R.M.  
TITLE Cloning of adeno-associated virus type 4 (AAV4) and generation of recombinant AAV4 particles  
JOURNAL J. Virol. 71 (9), 6823-6833 (1997)  
MEDLINE 97404695  
PUBMED 9261407  
REFERENCE 2 (bases 1 to 4767)  
AUTHORS Chiorini, J.A., Yang, L., Kotin, R.M. and Safer, B.  
TITLE Direct Submission  
JOURNAL Submitted (17-FEB-1997) MHB, NHLBI, NIH, Bld 10/7D18 10 Center Dr. MSC1654, Bethesda, MD 20982-1654, USA  
FEATURES  
source location/Qualifiers  
1. 4767  
/organism="adeno-associated virus 4"  
/isolate="ATCC VR-646"  
/db\_xref="taxon:57579"  
372..2243  
/codon\_start=1  
/product="rep"  
/protein\_id="AAC58044.1"  
/db\_xref="GI:2337938"  
/translation="MPGFYEIVLKVPSDLDEHLPGISDFSVNVAEKEMELPPDSMD  
LNLIEQAPLTVAEKIQREFLVWRRVSKAPEALFVQFEKGSYFHLILVETVSKS  
MNVGRVSOIKKLVTRIRYGRVEPQLPNWFAVTKRNGAGGNKVDDCYIPNYLPR  
TOPELOWMTNMDQYISACILAEKRLVAQHLTHSOTOENKONPNSDAPVIRS  
KTSARIMELVGLVDRGITSEKQWIDQOASISFNAASNSRQIKALDNASKIMSL  
TKTAPDYLVGQNPEDISSNRIRILNMNGYDPOYAAVFLGMAOKKFGKNTIWLFG  
PATGKTNIABAIHAHVPFYGCNVMTNENPFNDQVDMVIMWEEKMTAKVESAKA  
ILGSKRVVDQKCKSSAQIDPTPIVTSNTMCAVIDGNSTFEHQPLQDRMEKFEI  
TKRLEHDFGKVTKEVKDFFRWASDHTEVTHEFYVRKGARKRPAPNDADISEPKRA  
CPVSAQBSTSDAEAPVDYADRYQNKCSRHVGMNMLFPCRQERMNQNDICFTHGM  
DCAECFPVSESQPVSVVRKTYQKLCPIHIMGRABEVACSACELANVDLDDCMEQ"  
CDS  
2260..4464  
/codon\_start=1  
/product="capsid"  
/protein\_id="AAC58045.1"  
/db\_xref="GI:2337940"  
/translation="MTDGYLPDWLEDNLSEGVREWWALQPGAPRKANQOHQDNARGL  
VLPGYKYLGPNGLDKGEPVNAADAALEHDKAYDQOLKAGDNPLKYNHADAEFQQR  
LQGDTSFGGNLGRAVFOAKKRVLEPLGLVEQAGETAPGKRPLIESPQPPDSSTGIG  
KGKOPAKKLVFEDETGAGDPREGSTGAMSDDSEMRAGAAGGAGGAGADGVNA  
SGDWCHDSSTWSEGHVTTSTRITWLTPTNNHLYKRLGESLQSTNGFSTPWGYPDN  
RFHCHFSRDMWRLLINNMWGRPKAMRVKIFNIOVKEVTSNGETVANNLTSTVOIF  
ADSSYELPYMDAGQEGSLPPFPNDVPMVPOYGCGLVGNTSQOQTDNAPCYCLEYF  
PSQMLRTGNNEIITYSEKVPFHSVMYHQSOLDRLMNPILDQYLMGLQSTTTGTLNA  
GTATNFTKLRPNTFSNPKKNWLPGPSIKQGGSKTANQNYKIPATGSDSLIKYETHS  
TLDGRMSALTGPPMATAGPADSKNSQLIFAGPKQNGNTATVPGTLIFTSEELAA  
TNATIDTMWGNLPGDQSNLPTVDRLLTALGAVPGMWQNRDIYYQGPWAKIPIHTD

GHFNPSPILGGFLKHPPEQIFIKNTPVPANPATFESSTPVNSFITQYSTGVSVQID  
WEIQKERSKRMNPEVQFTSNYGQNSLWAPDAKYTEPRAIGTRYLTHHL"  
BASE COUNT 1137 a 1405 c 1333 g 892 t  
ORIGIN  
Query Match 55.1%; Score 2598.6; DB 14; Length 4767;  
Best local Similarity 74.6%; Pred. No. 0;  
Matches 3584; Conservative 0; Mismatches 1094; Indels 129; Gaps 20;  
Qy 1 TTGCCACTCCCTCTCTGCGCGCTGCTGCTCGGTGGGGCTGCGGACCAAGGTCGC 60  
Db 1 TTGGCACTCCCTCTATGCGCGCTGCTCACTCACTCGCCCTGAGACCAAGGTCGC 60  
Qy 61 AGACGGCAGAGCTCTGCTCTGCGGCCCCACCGAGCGAGCGCGCAGAGGAGTG 120  
Db 61 AGACTGCGGCTCTGCGCGGAGCGCGAGTGAGTGAGCGAGCGCATAGAGGAGTG 120  
Qy 121 GGCACTCATCACTAGG-----GTAAATCGGAAGCGCTCCAC 161  
Db 121 GCCAATCATCATCTAGGTTGCCCACTGACGTCAATGTGACGCTTAGGGAG 180  
Qy 162 GCTGCCGGCTCAGCGCTGACGTAATTAAGTCATAGGGAGTGTCTGTA-----T 213  
Db 181 GTCCCTGATTAAGACGTACAGTACGTGTCGATTTCCGGAGCGTACGGAGCGCATACC 240  
Qy 214 TAGCTGTCAGTGA-----GTGCTTTGGCAGATTGCGACACACCGTGGCCAT 263  
Db 241 AAGCTGCACGTACAGACGACGAGTGTGCTTTGGACAGTTGCGACACCATGTGTGAG 300  
Qy 264 TTAGGATATATATGCGCGAGTGAGGAGC-AGGATCTCCATTGTAACCGGAATTGAA 322  
Db 301 G-AGGATATATAACCGGAGTGAGCGACGAGAGTCCATTGCGCGGAATTGAA 359  
Qy 323 CGAGCAGAGCATGCGCGGCTTCTAGAGATCGTATGATCAAGGTGCGGAGCGACTGAC 382  
Db 360 CGAGCAGAGCATGCGGGGTTCTAGAGATCGTGTGAGGTGCCAGCGACTGAG 419  
Qy 383 GAGCACTGCGCGGCAATTCTGACTGTTTGAGCTGGGTGCCGAGAAGATGGAG 442  
Db 420 GAGCACTGCGCGGCAATTCTGACTTTTGTGAGCTGGGTGCCGAGAAGATGGAG 479  
Qy 443 CTGCCCCCGGATTTGACATGATGTAATCTGATTGAGCAGGACCCCTGACCGTGCC 502  
Db 480 CTGCGCGGATTTGACATGACTTGAATCTGATTGAGCAGGACCCCTGACCGTGCC 539  
Qy 503 GAGAGCTGACGCGGACTTCTGCTCAATGCGCGCGCGGTAGTAAGCCCGGAGGCC 562  
Db 540 GAAAAGCTGCAACGCGAGTCTGTGAGTGCGCGCGCGGTAGTAAGCCCGGAGGCC 599  
Qy 563 CTCTCTTGTGATTCGAGAAAGGCGAGTCTCACTTCCACTCATATTCTGTGTGAG 622  
Db 600 CTCTCTTGTGATTCGAGAAAGGCGAGTCTCACTTCCACTCATATTCTGTGTGAG 659  
Qy 623 ACCACGGGGTCAATCCATGCTGCGCGCTTCTGAGTCAAGTTAGGACAAGCTG 682  
Db 660 ACCGTGGGCTCAATCCATGCTGCGCGCGCTGAGTCAAGTTAGGACAAGCTG 719  
Qy 683 GTGCAAGCATCTACCGCGGATCGAGCCGACCTGCCCACTGCTGCGGCTGACCAAG 742  
Db 720 GTGACCGCATCTACCGCGGATCGAGCCGACCTGCCCACTGCTGCGGCTGACCAAG 779  
Qy 743 ACGGTAATGCGCGCGGAGGGGGGACAAGGTGTGAGCAGTGTCTCATCCCACTAC 802  
Db 780 ACGGTAATGCGCGCGGAGGGGGGACAAGGTGTGAGCAGTGTCTCATCCCACTAC 839  
Qy 803 CTCTGCGCAAGACTCAGCCCGAGCTGAGTGGCGGTGACTTAATGAGAGATATATA 862  
Db 840 CTGCTCCCAAGACCCAGCCGAGCTCAGTGGCGGTGACTTAATGAGAGATATATA 899  
Qy 863 AGGCGCTTTGAACCTGCGGAGCGCAAGCGCTGCTGCGGAGCAGCAGTCAACCGTC 922  
Db 900 AGGCGCTTTGAATCTCGCGGAGCGTAACGCGCTGCTGCGGAGCAGCAGTCAACCGTC 959







Db 3084 CAACCGCTTCCACTGCTCACTTCTCACCACGTGACTGGACGCACTCATCAACAACACTG 3143  
QY 3137 GGGATTCGGGCGGAGAGAGACTCAACTTCAAACTCTTCAACATCCAAAGTCAAGAGGTCCAC 3196  
Db 3144 GGGCATGCGAGCCCAAGCCATGCGGTCAAAATCTTCAACATCCAGGTCAAGAGGTCCAC 3203  
QY 3197 GACGAATGATGGCTCAACAACCATCGCTAATACTTACCAGCAGCGGTCAAGTCTTCTC 3256  
Db 3204 GACGTGCAACGGCGAGACACAGGTGGCTAATACTTACCAGCAGCGGTCAAGTCTTCTC 3263  
QY 3257 GGACTCGAGTACCAAGCTTCCGTACGTCTCGGCTCTGGCAGCAGGCTGCTCCCTCC 3316  
Db 3264 GGACTCGTCTGACGAATGCGGTACGTGATGATGCGGCTCAAGAGGGCAGCCTGCTCC 3323  
QY 3317 GTTCCCGGCGGAGCTGTTCATGATTCGCAATACGGCTAC-----CTGACGCTCAA 3367  
Db 3324 TTTTCCCAACGAGCTCTTATGTGCCCCAGTACGGCTACTGTGACTGTGACCGGCAA 3383  
QY 3368 CAATGGCAGCCCAAGCCGTGGGAGCTTCAATCCTTTTACTGCTGGAATATTTCCCTCTCA 3427  
Db 3384 CACTTCGACGAACAAGACTGACAGAATATGCTTCACTGCTGAGTACTTTCCTTCCGA 3443  
QY 3428 GATGCTGAGAACGGGCAACAATTTACTTCACTTCACTGACCTTTGAGAGTGCCTTTCCA 3487  
Db 3444 GATGCTGCGGACTGGCAACAATTTGAATATGTAAGTTTGAAGAGTGCCTTTCCA 3503  
QY 3488 CAGCAGTACGCGCAGCAGCAGAGCCTGAGCGGCTGATGATCCTTCATCGAACCAATA 3547  
Db 3504 CTCGATGTACGCGCAGCAGCAGAGCCTGAGCGGCTGATGAACTCTCATCGAACCAATA 3563  
QY 3548 CCTGTATTACTGACAGAACAAATCAATCAGTCCGGA---AGTCCCAAAACAAGACTT 3604  
Db 3564 CCTGTGGGAGTGTCAATCGAACCAACCGGAACCACTGATGATGCGGAGCTGCCACAC 3623  
QY 3605 GCTGTTTACGCGGTGCTCTCCAGCTGCGATGTCTGTTCAGCCCAAAACTGCGTACTG 3664  
Db 3624 CAACTTTACCAAGCTGCGGCTTCACTTCACTTTTCACTTTAAAGAACTGCGTCCCG 3683  
QY 3665 ACCCTGTTATCGGAGCAGCGCGCTTCTTAAACAAACAAAGACAAACAAAGCAATTT 3724  
Db 3684 GCCTTCAATCAAGAGCAGAGGCTTCTCAAGACTGCAATCAAAACTTCAAGATCCCTGC 3743  
QY 3725 TACCTGAGCTGTGCTTCAAAATATAAC-----CTCAATGGGCGTGAATC 3769  
Db 3744 CACCGGTCAGACAGTCTCATCAATAACAGACGACAGCACTGTGACGGAAGATGAG 3803  
QY 3770 CATCATCAACCTGCGCACTGCTATGCTTACACAAGAGCAGAAAGCAAGTTCTTCC 3829  
Db 3804 TGCCCTGACCCCGGAGCTTCCATATGCGCAGCGCTGAGCAGCAAGTTCCAG-- 3860  
QY 3830 CATGAGCGGTTCATGATTTTGGAAAAGAGCGCCGAGCTTCAACACTGCATTGGA 3889  
Db 3861 CAACAGCCAGCTCATTTTGGGGGCTTAAACAGAAACGGAACAACGGAACCGTACC 3920  
QY 3890 CAATGTATGATTAACAAGAGAGAAATTAAGCCACTAACCTGTGAGCCACGAAAG 3949  
Db 3921 GACTCTGATCTTCACTCTGAGAGAGAGCTGGAGCCACCAACGCCAGATACGGACAT 3980  
QY 3950 ATTTGGGACCGTGGCAGTCAATTTCCAGAGCAGACAGACAGACCTGCGACCGGAGATGT 4009  
Db 3981 GTGGGGCAACTACTGCGGCTGACAGAGCAACAGCAACCTGCGACCGTGGACAGACT 4040  
QY 4010 GCATGCTATGGAGCATTAACCTGGCATGGTGGCAAGATAGAGACGTACTGACAGG 4069  
Db 4041 GACAGCCTTGGGAGCGCTGCTGGAATGTGTGGCAAAACAGAGACATTTACTACAGG 4100  
QY 4070 TCCCATTTGGGCAAAATTCCTCACACAGATGACATTTCAACCGCTCTCTTATGGG 4129  
Db 4101 TCCCATTTGGGCAAGATTCCTCATACCGATGACATTTCAACCGCTCATGATGG 4160  
QY 4130 CGGCTTTGAGTCAAGAACCCGCTCTCAGATCCTCATCAAAAACAGCCTGTCTCTGC 4189  
Db 4161 TGGGTTTGGGCTGAACAACCCGCTCTCTCAAAATTTTATCAAGAACACCCGCTACCTGC 4220

QY 4190 GAATCCTCCGGCGGAGTTTTCAGCTACAAAGTTTGCTTCAATTCATCAACCAATACTCCAC 4249  
Db 4221 GAATCCTGCAACGACCTTCACTCTACTCCGGTAAACTCTTCAATTAAGTACAGACAC 4280  
QY 4250 AGGACAAAGTGTGTGAAATTTGAATGGAGCTGCAGAAAGAAAACAGCAAGCGCTGGA 4309  
Db 4281 TGGCAGGTGTGGTGAGATTTGACTGGAGATCCAGAGAGCGGCTCAAAACGCTGGA 4340  
QY 4310 TCCCGAAGTGCAGTACACATCCAAATTATGCAAAATCTGCCAAAGTTGATTTACTGTGA 4369  
Db 4341 CCCGAGGTCCAGTTTACCTCCAACTACGACAGCAAAACTCTGTGTGGCTCCGA 4400  
QY 4370 CAACAATGACTTATATGAGCTGCGCCCATGAGCAGCCGTTACCTTACCCGCTCCCT 4429  
Db 4401 TGCGGCTGGGAAATACACTGAGCCTAGGCTATCGGTACCCGCTACCTCACCCACCACT 4460  
QY 4430 GTAATGCTGTATATCAATAAACCGGTTGATTCGTTTCACTTGAATTTGCTCTCC-TG 4488  
Db 4461 GTAATACCTGTATATCAATAAACCGGTTTATTCGTTTCACTTGAATTTGCTCTCCGTG 4520  
QY 4489 TCCTTCTTATCTTATC-GGTACCAATGTTATAGCTTACACATTAATG-----CTTG 4541  
Db 4521 TCCTTCTTATCTTATCTGTTTCACTGCTTCACTGCTTACGCTACATTAAGCAGCGCTG 4580  
QY 4542 TTGCGCTTGGCATTAAGACTTACG-----TCATCGGTTTACCCCTAGTATG 4591  
Db 4581 TTGCGCTTGGCGTTTACAACTGCGGTTTATCAGTAACTTCTGCGCAAAACAGATGATGG 4640  
QY 4592 AGTTGCCACTCCCTCTGTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4651  
Db 4641 AGTTGCCACTTATGCTATGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4700  
QY 4652 GCAGACGAGAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4711  
Db 4701 CCAGACTGCGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4760  
QY 4712 TGGGCAA 4718  
Db 4761 TGGGCAA 4767

Search completed: July 20, 2003, 13:19:53  
Job time : 12464.8 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2003, 07:53:46 ; Search time 32.1207 Seconds  
(without alignments)  
10025.800 Million cell updates/sec

Title: US-09-807-802A-1\_COPY\_4576\_4718

Perfect score: 143  
Sequence: 1 ttaccctagtgtgagctt.....cgacagagggagtgggca 143

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_101002:\*

1:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	143	100.0	4718	21	AAD00772 Adeno-associated v
2	125	87.4	143	21	AAD00781 Adeno-associated v
3	121.8	85.2	4683	22	AAF23749 AAV6 DNA sequence
4	119	83.2	143	21	AAD00781 Adeno-associated v
5	119	83.2	4718	21	AAD00772 Adeno-associated v
6	114.8	80.3	4683	22	AAF23749 AAV6 DNA sequence
7	107.4	75.1	145	18	AAT63408 Inverted terminal
8	107.4	75.1	145	24	ABA02990 Adeno-associated v
9	107.4	75.1	165	15	AAQ66769 Double-D ITR termi

10	107.4	75.1	165	18	AAT49462 Adeno associated v
11	107.4	75.1	165	22	AAD03535 Inverted terminal
12	107.4	75.1	192	16	AAT03384 Right-hand Inverte
13	107.4	75.1	272	21	AAZ47166 Head-to-tail ITR s
14	107.4	75.1	272	22	AAD03534 AAV circular inter
15	107.4	75.1	4675	22	AAH41481 Adeno-associated v
16	107.4	75.1	4675	22	AAF89931 Nucleotide sequenc
17	107.4	75.1	4675	22	ABA02989 Adeno-associated v
18	107.4	75.1	4679	22	AAI66974 Adeno-associated v
19	107.4	75.1	4679	22	AAF23750 AAV2 DNA sequence
20	107.4	75.1	4680	17	AAT09008 Wild-type adeno-as
21	107.4	75.1	8698	22	AAH26325 Wild-type adeno-as
22	105.8	74.0	189	24	AAI71057 AAV terminal repea
23	105.8	74.0	191	24	ABL58791 AAV 5'ITR from AAV
24	105.8	74.0	272	21	AAZ47164 Head-to-tail ITR s
25	105.8	74.0	272	22	AAD03532 AAV circular inter
26	105.4	73.7	272	21	AAA10404 Head-to-tail ITR s
27	102.6	71.7	272	21	AAZ47166 Head-to-tail ITR s
28	102.6	71.7	272	22	AAD03534 AAV circular inter
29	101	70.6	174	24	ABL58793 AAV 3'ITR from AAV
30	101	70.6	183	21	AAA10405 Human adeno-associ
31	99.4	69.5	145	14	AAQ41448 AAV2 inverted term
32	99.4	69.5	145	16	AAT03385 Strict inverted te
33	99.4	69.5	145	20	AAH34295 Adeno-associated v
34	99.4	69.5	165	15	AAQ66769 Double-D ITR termi
35	99.4	69.5	165	18	AAT49462 Adeno associated v
36	99.4	69.5	165	22	AAD03535 Inverted terminal
37	99.4	69.5	955	24	AAD37254 Adeno-associated v
38	99.4	69.5	955	24	AAD37254 Adeno-associated v
39	99.4	69.5	987	24	AAD37261 Adeno-associated v
40	99.4	69.5	987	24	AAD37260 Adeno-associated v
41	99.4	69.5	4414	24	AAD37260 Adeno-associated v
42	99.4	69.5	4414	24	AAD37260 Adeno-associated v
43	99.4	69.5	4476	24	AAD37259 Adeno-associated v
44	99.4	69.5	4476	24	AAD37259 Adeno-associated v
45	99.4	69.5	4498	24	AAD37258 Adeno-associated v

## ALIGNMENTS

RESULT 1	
AAD00772	
ID	AAD00772 standard; DNA; 4718 BP.
XX	
AC	AAD00772;
XX	
DT	08-SEP-2000 (first entry)
XX	
DE	Adeno-associated virus serotype 1 DNA.
XX	
KW	Adeno-associated virus serotype 1; AAV-1; rep protein; capsid protein;
KW	cap protein; recombinant viral vector; gene delivery; gene therapy;
KW	vaccine; transgene; ss.
XX	
OS	Adeno associated virus serotype 1.
XX	
FH	Key
FT	repeat_unit
FT	Location/Qualifiers
FT	1.143
FT	/tag= a
FT	/label= 5' ITR
FT	/note= "Inverted terminal repeat which is capable
FT	of forming T-shaped hairpin structure"
FT	89.110
FT	/tag= b
FT	/bound moiety= "Rep protein"
FT	124.125
FT	/tag= c
FT	/note= "Terminal resolute site (TRS)"
FT	219.226
FT	/tag= d
FT	/bound moiety= "USF"
FT	/note= "E box"
FT	

FT promoter 236..299  
/tag= e  
/label= p5\_promoter  
FT protein\_bind 237..245  
/tag= f  
/bound\_moiety= "Y1 factor"  
FT TATA\_signal 270..275  
/tag= g  
/label= p5\_TATA-Box  
FT misc\_feature 299..306  
/tag= h  
/note= "Y1/p5 RNA"  
FT CDS 335..2206  
/tag= i  
/product= "Rep 78"  
/function= "regulates replication and integration  
of AAV DNA into host cell's chromosome"  
FT CDS 335..2272  
/tag= j  
/product= "Rep 68"  
/function= "regulates replication and integration  
of AAV DNA into host cell's chromosome"  
FT intron 1924..2220  
/tag= k  
/note= "This region interrupts the coding sequence  
of Rep 68 and Rep 40"  
FT TATA\_signal 857..862  
/tag= l  
/label= p19\_TATA\_Box  
FT misc\_feature 882..883  
/tag= m  
/note= "p19 RNA"  
FT CDS 1007..2206  
/tag= n  
/product= "Rep 52"  
/function= "regulates replication and integration  
of AAV DNA into host cell's chromosome"  
FT CDS 1007..2272  
/tag= o  
/product= "Rep 40"  
/function= "regulates replication and integration  
of AAV DNA into host cell's chromosome"  
FT TATA\_signal 1836..1841  
/tag= p  
/label= p40\_TATA-Box  
FT misc\_feature 1875..1876  
/tag= q  
/note= "P40 RNA"  
FT CDS 2223..4433  
/tag= r  
/product= "VP1 protein"  
/note= "Capsid protein"  
FT CDS 2634..4433  
/tag= s  
/product= "VP2 protein"  
/note= "Capsid protein"  
FT CDS 2829..4433  
/tag= t  
/product= "VP3 protein"  
/note= "Capsid protein"  
FT polyA\_signal 4447..4452  
/tag= u  
FT repeat\_unit 4576..4718  
/tag= v  
/label= 3' ITR  
/note= "Inverted terminal repeat which is capable  
of forming T-shaped hairpin structure"  
XX WO200028061-A2  
PN  
XX

PD 18-MAY-2000.  
XX  
XX 02-NOV-1999; 99WO-US25694.  
XX  
XX 05-NOV-1998; 98US-0107114.  
XX  
XX (TYPE-) UNIV PENNSYLVANIA.  
XX  
XX Wilson JM, Xiao W;  
XX  
XX MPI; 2000-376571/32.  
DR P-PSDB; AAY71161, AAY71164, AAY71165, AAY71166, AAY71167, AAY71168,  
DR AAY71169.  
XX  
XX Novel adeno-associated virus serotype 1 polynucleotide useful for  
PT preparation of medicament for delivery of a transgene to a host  
XX  
XX Claim 1; Fig 1; 108bp; English.  
PS  
XX  
XX The present sequence is an adeno-associated virus serotype 1 (AAV-1)  
CC DNA characterised by two inverted terminal repeats (ITR) and open  
CC reading frames for rep and capsid (cap) proteins. The rep reading frame  
CC encodes four proteins, Rep 78, Rep 68, Rep 52 and Rep 40, while the cap  
CC reading frame encodes three structural proteins, VP1, VP2 and VP3.  
CC The AAV-1 sequence or its fragments particularly ITRs, rep and cap  
CC coding regions, are useful in production of recombinant viral vectors  
CC for gene delivery. These vectors can be used as gene therapy  
CC vectors, vaccine vectors or antisense delivery vectors. The AAV-1 does  
CC not induce the formation of neutralising antibodies specific to any  
CC serotype of AAV hence is useful for transforming host cells, and in the  
CC preparation of a medicament for the delivery of transgene to a host.  
XX  
XX Sequence 4718 BP; 1121 A; 1393 C; 1273 G; 931 T; 0 other;  
SQ  
Query Match 100.0%; Score 143; DB 21; Length 4718;  
Best Local Similarity 100.0%; Pred. No. 1.8e-28;  
Matches 143; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TTACCCCTAGTATGAGTGTGCCACTCCTCTCTGCGCGCTCGCTCGGTGGGCC 60  
DB 4576 TTACCCCTAGTATGAGTGTGCCACTCCTCTCTGCGCGCTCGCTCGGTGGGCC 4635  
QY 61 TGGGACCAAGGTCCGACAGCGGAGCTGTGCTGCGCGGCCCAACGAGCGCA 120  
DB 4636 TGGGACCAAGGTCCGACAGCGGAGCTGTGCTGCGCGGCCCAACGAGCGCA 4695  
QY 121 GCGCGCAGAGCGAGTGGGCA 143  
DB 4696 GCGCGCAGAGCGAGTGGGCA 4718  
RESULT 2  
AAD00781  
ID AAD00781 standard; DNA; 143 BP.  
XX  
AC AAD00781;  
XX  
DT 08-SEP-2000 (first entry)  
XX  
DE Adeno-associated virus serotype 1 5' inverted terminal repeat.  
XX  
KW Adeno-associated virus serotype 1; AAV-1; rep protein; capsid protein;  
KW cap protein; recombinant viral vector; gene delivery; gene therapy;  
KW vaccine; transgene; inverted terminal repeat; ITR; ss.  
XX  
OS Adeno associated virus serotype 1.  
XX  
XX  
FH Key Location/Qualifiers  
FT misc\_binding 1..41  
FT /tag= a  
FT /bound\_moiety= "binds to nucleotides 85..125"  
FT stem\_loop 42..62  
FT /tag= b  
FT



```
FT stem_loop 64..84
FT misc_binding /+tag= c
FT 85..125
FT /+tag= d
FT /bound_moiety= "binds to nucleotides 41..1"
XX PN WO200028061-A2.
XX PD 18-MAY-2000.
XX PF 02-NOV-1999; 99WO-US25694.
XX PR 05-NOV-1998; 98US-0107114.
XX PA (UYPE-) UNIV PENNSYLVANIA.
XX PI Wilson JM, Xiao W;
XX DR WPI; 2000-376571/32.
XX PT Novel adeno-associated virus serotype 1 polynucleotide useful for
XX preparation of medicament for delivery of a transgene to a host -
XX PS Example 2; Fig 2, 108pp; English.
XX CC The patent discloses an adeno-associated virus serotype 1 (AAV-1) DNA
XX which is characterised by two inverted terminal repeats (ITR) and open
XX reading frames for rep and capsid (cap) proteins. The rep reading frame
XX encodes four proteins, Rep 78, Rep 68, Rep 52 and Rep 40, while the cap
XX reading frame encodes three structural proteins, VP1, VP2 and VP3.
XX The AAV-1 sequence or its fragments particularly ITRs, rep and cap
XX coding regions, are useful in production of recombinant viral vectors
XX for gene delivery. These vectors can be used as gene therapy
XX vectors, vaccine vectors or antisense delivery vectors. The AAV-1 does
XX not induce the formation of neutralising antibodies specific to any
XX serotype of AAV hence is useful for transforming host cells, and in the
XX preparation of a medicament for the delivery of transgene to a host.
XX The present sequence is a 5' ITR of AAV-1 DNA which is useful in the
XX production of recombinant viral vector. The ITR forms a T-shaped hairpin
XX structure.
XX SQ Sequence 143 BP; 24 A; 50 C; 48 G; 21 T; 0 other;
Query Match 87.4%; Score 125; DB 21; Length 143;
Best Local Similarity 100.0%; Pred. No. 7.6e-24;
Matches 125; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 19 TTGCCCACTCCCTCTCTGCGGCTCGCTCGGTGGGGCTGCGGACCAAGTCCGC 78
Db 1 TTGCCCACTCCCTCTCTGCGGCTCGCTCGGTGGGGCTGCGGACCAAGTCCGC 60
OY 79 AGACGGCAGAGCTCTGCTCTGCCGCGCCCAACGAGCGAGCGCGCAGAGAGGAGTG 138
Db 61 AGACGGCAGAGCTCTGCTCTGCCGCGCCCAACGAGCGAGCGCGCAGAGAGGAGTG 120
OY 139 GGCAA 143
Db 121 GGCAA 125
RESULT 3
AAF23749/c
ID AAF23749 standard; DNA; 4683 BP.
XX AC AAF23749;
XX DT 28-MAR-2001 (first entry)
XX DE AAV6 DNA sequence.
XX KW AAV6; gene therapy; AAV viral vector; cystic fibrosis; cancer; AIDS;
XX atherosclerosis; sickle cell anaemia; thalassaemia;
KW blood clotting disorder; diabetes; ss.
```

```
XX OS Adeno associated virus.
XX PN US6156303-A.
XX PD 05-DEC-2000.
XX PF 11-JUN-1997; 97US-0873168.
XX PR 11-JUN-1997; 97US-0873168.
XX PA (UNIW ) UNIV WASHINGTON.
XX PI Russell DW, Rutledge EA;
XX DR WPI; 2001-060164/07.
XX PT Adeno-associated virus serotype 6 and viral vector derived from it for
XX gene therapy of cystic fibrosis, cancer, acquired immunodeficiency
XX syndrome, sickle cell anemia, thalassemia and diabetes -
XX PS Claim 1; Fig 1; 50pp; English.
XX CC The present invention relates to adeno-associated virus serotypes. The
XX present sequence is the DNA sequence of one such serotype (AAV6). AAV6
XX can be used to construct AAV viral vectors for use in gene therapy for a
XX range of disorders: cystic fibrosis, cancer, AIDS, atherosclerosis,
XX sickle cell anaemia, thalassaemia, blood clotting disorders and diabetes.
XX The AAV viral vectors have increased transduction efficiency of a
XX particular host cell as the AAV virion containing the AAV vector genome
XX can be modified to express a capsid protein of an AAV serotype that
XX transduces the selected host cell.
XX SQ Sequence 4683 BP; 1114 A; 1363 C; 1277 G; 929 T; 0 other;
Query Match 85.2%; Score 121.8; DB 22; Length 4683;
Best Local Similarity 98.4%; Pred. No. 6.1e-23;
Matches 123; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 19 TTGCCCACTCCCTCTCTGCGGCTCGCTCGGTGGGGCTGCGGACCAAGTCCGC 78
Db 4683 TTGCCCACTCCCTCTATAGCGGCTCGCTCGGTGGGGCTGCGGACCAAGTCCGC 4624
OY 79 AGACGGCAGAGCTCTGCTCTGCCGCGCCCAACGAGCGAGCGCGCAGAGAGGAGTG 138
Db 4623 AGACGGCAGAGCTCTGCTCTGCCGCGCCCAACGAGCGAGCGCGCATAGAGGAGTG 4564
OY 139 GGCAA 143
Db 4563 GGCAA 4559
RESULT 4
AAD00781/c
ID AAD00781 standard; DNA; 143 BP.
XX AC AAD00781;
XX DT 08-SEP-2000 (first entry)
XX DE Adeno-associated virus serotype 1 5' inverted terminal repeat.
XX KW Adeno-associated virus serotype 1; AAV-1; rep protein; capsid protein;
XX cap protein; recombinant viral vector; gene delivery; gene therapy;
XX vaccine; transgene; inverted terminal repeat; ITR; ss.
XX OS Adeno associated virus serotype 1.
XX FH Key Location/Qualifiers
FH misc_binding 1..41
FT /+tag= a
FT /bound_moiety= "binds to nucleotides 85..125"
FT stem_loop 42..62
```

```
FT      stem_loop      /*tag= b
FT      misc_binding    /*tag= c
FT      misc_binding    85..125
FT      /*tag= d
FT      /bound_moiety= "binds to nucleotides 41..1"
XX      WO200028061-A2.
XX      18-MAY-2000.
XX
XX      02-NOV-1999;    99WO-US25694.
XX
XX      05-NOV-1998;    98US-0107114.
XX
XX      (UYPE-) UNIV PENNSYLVANIA.
XX
XX      Wilson JM, Xiao W;
XX
XX      WPI, 2000-376571/32.
XX
XX      Novel adeno-associated virus serotype 1 polynucleotide useful for
XX      preparation of medicament for delivery of a transgene to a host
XX
XX      Example 2; Fig 2; 108pp; English.
PS
CC      The patent discloses an adeno-associated virus serotype 1 (AAV-1) DNA
CC      which is characterised by two inverted terminal repeats (ITR) and open
CC      reading frames for rep and capsid (cap) proteins. The rep reading frame
CC      encodes four proteins, Rep 78, Rep 68, Rep 52 and Rep 40, while the cap
CC      reading frame encodes three structural proteins, VP1, VP2 and VP3.
CC      The AAV-1 sequence or its fragments particularly ITRs, rep and cap
CC      coding regions, are useful in production of recombinant viral vectors
CC      for gene delivery. These vectors can be used as gene therapy
CC      vectors, vaccine vectors or antisense delivery vectors. The AAV-1 does
CC      not induce the formation of neutralising antibodies specific to any
CC      serotype of AAV hence is useful for transforming host cells, and in the
CC      preparation of a medicament for the delivery of transgene to a host.
CC      The present sequence is a 5' ITR of AAV-1 DNA which is useful in the
CC      production of recombinant viral vector. The ITR forms a T-shaped hairpin
CC      structure.
CC
XX      Sequence 143 BP; 24 A; 50 C; 48 G; 21 T; 0 other;
SQ
Query Match      83.2%; Score 119; DB 21; Length 143;
Best Local Similarity 89.5%; Pred. No. 2.8e-22;
Matches 128; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
OY      1 TTACCCCTAGTGTGAGTGTGCCCACTCCCTCTGTGCGCGCTGCTGCTGGTGGGCC 60
      |||||||
      143 TTACCCCTAGTGTGAGTGTGCCCACTCCCTCTGTGCGCGCTGCTGCTGGTGGGCC 84
OY      61 TGGGACCAAGGTCGCGACAGCGGACAGAGCTCTGCTCTGCCGCGCCACCGAGGAGCGA 120
      |||||||
      83 GGCAGAGCAGAGCTGTGCGCTGTGCGACCTTGTGTCGACGCGCCACCGAGGAGCGA 24
OY      121 GCGCGCAGAGAGGAGTGGGCAA 143
      |||||||
      23 GCGCGCAGAGAGGAGTGGGCAA 1
Db
```

```
KW      vaccine; transgene; ss.
XX
XX      Adeno associated virus serotype 1.
OS
XX      Key
FH      repeat_unit    Location/Qualifiers
FT      1..143
FT      /*tag= a
FT      /label= 5' ITR
FT      /note= "Inverted terminal repeat which is capable
FT      of forming T-shaped hairpin structure"
FT      89..110
FT      /*tag= b
FT      /bound_moiety= "Rep protein"
FT      124..125
FT      /*tag= c
FT      /note= "Terminal resolute site (TRS)"
FT      219..226
FT      /*tag= d
FT      /bound_moiety= "USF"
FT      /note= "E box"
FT      236..299
FT      /*tag= e
FT      /label= P5_promoter
FT      237..245
FT      /*tag= f
FT      /bound_moiety= "YY1 factor"
FT      270..275
FT      /*tag= g
FT      /label= P5_TATA-Box
FT      299..306
FT      /*tag= h
FT      /note= "YY1/p5 RNA"
FT      335..2206
FT      /*tag= i
FT      /product= "Rep 78"
FT      /function= "regulates replication and integration
FT      of AAV DNA into host cell's chromosome"
FT      335..2272
FT      /*tag= j
FT      /product= "Rep 68"
FT      /function= "regulates replication and integration
FT      of AAV DNA into host cell's chromosome"
FT      1924..2220
FT      /*tag= k
FT      /note= "This region interrupts the coding sequence
FT      of Rep 68 and Rep 40"
FT      857..862
FT      /*tag= l
FT      /label= P19_TATA_Box
FT      882..883
FT      /*tag= m
FT      /note= "P19 RNA"
FT      1007..2206
FT      /*tag= n
FT      /product= "Rep 52"
FT      /function= "regulates replication and integration
FT      of AAV DNA into host cell's chromosome"
FT      1007..2272
FT      /*tag= o
FT      /product= "Rep 40"
FT      /function= "regulates replication and integration
FT      of AAV DNA into host cell's chromosome"
FT      1836..1841
FT      /*tag= p
FT      /label= P40_TATA-BOX
FT      1875..1876
FT      /*tag= q
FT      /note= "P40 RNA"
FT      2223..4433
FT      /*tag= r
FT      /product= "VP1 protein"
```

FT		/note= "Capsid protein"
FT	CDS	2634..4433
FT		/+tag= s
FT		/product= "VP2 protein"
FT		/note= "Capsid protein"
FT		/partial
FT	CDS	2829..4433
FT		/+tag= t
FT		/product= "VP3 protein"
FT		/note= "Capsid protein"
FT	polyA_signal	4447..4452
FT		/+tag= u
FT	repeat_unit	4576..4718
FT		/+tag= v
FT		/label= 3'_ITR
FT		/note= "Inverted terminal repeat which is capable
FT		of forming T-shaped hairpin structure"

WO200028061-A2

18-MAY-2000

02-NOV-1999; 99WO-US25694.

05-NOV-1998; 98US-0107114.

(UNIVERSITY OF) PENNSYLVANIA

Wilson JM, Xiao W;

WPI; 2000-376571/32

AA771169.

Novel adeno-associated virus serotype 1 polynucleotide useful for

Claim 1; Fig 1; 108pp; English.

CC The present sequence is an adeno-associated virus serotype 1 (AAV-1)  
CC DNA characterised by two inverted terminal repeats (ITR) and open  
CC reading frames for rep and capsid (cap) proteins. The rep reading frame  
CC encodes four proteins, Rep 78, Rep 68, Rep 52 and Rep 40, while the cap  
CC reading frame encodes three structural proteins, VP1, VP2 and VP3.  
CC The AAV-1 sequence or its fragments particularly ITRs, rep and cap  
CC coding regions, are useful in production of recombinant viral vectors  
CC for gene delivery. These vectors can be used as gene therapy  
CC vectors, vaccine vectors or antisense delivery vectors. The AAV-1 does  
CC not induce the formation of neutralising antibodies specific to any  
CC serotype of AAV hence is useful for transforming host cells, and in the  
CC preparation of a medicament for the delivery of transgene to a host.

Sequence 4718 BP; 1121 A; 1393 C; 1273 G; 931 T; 0 other;

Query Match	83.2%;	Score 119;	DB 21;	Length 4718;
Best Local Similarity	89.5%;	Pred. No. 3.3e-22;		
Matches 128; Conservative	0;	Mismatches 15;	Indels 0;	Gaps 0;

QY 1 TTACCCCTAGTGTGAGTTGCCCACTCCCTCTCTGCGGCGCTCGCTCGGTGGGCGC 60  
|||  
Db 143 TTACCCCTAGTGTGAGTTGCCCACTCCCTCTCTGCGGCGCTCGCTCGGTGGGCGC 84

Qy 61 TCCGGAACAAAGSTCCGCAGACGGCAGACTCTGCTTGCGGGCCCCACCAGCGACCGA 120  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 83 GCGACAGCAGAGCTTGTCCGTCTGCCGACCTTTGGTCCGCAGGCCCCACCAGCGACCGCA 24

QY	121	GCGCGCAGAGGGAGTGGCAA	143
Db	23	GCGCGCAGAGGGAGTGGCAA	1

RESULT 6  
AAF23749

ID	AAF23749	standard; DNA; 4683 BP.

AC . AAF23749

DT 28-MAR-2001 (First entry)

AAV6 DNA sequence.

KW AAV6; gene therapy; AAV viral vector; cystic fibrosis; cancer; AIDS  
 KW atherosclerosis; sickle cell anaemia; thalassaemia;  
 KW blood clotting disorder; diabetes; ss.

OS Adeno associated virus.

PN US6156303-A

PD 05-DEC-2000

PF 11-JUN-1997; 97US-0873168

PR 11-JUN-1997; 97US-0873168

PA (UNIW ) UNIV WASHINGTON

PI Russell DW, Rutledge EA;

DR WPI; 2001-060164/07

PT Adeno-associated virus serotype 6 and viral vector derived from it for

PT syndrome, sickle cell anemia, thalassemia and diabetes -

PS Claim 1; Fig 1; 50pp; English.

CC The present invention relates to adeno-associated virus serotypes. The  
CC present sequence is the DNA sequence of one such serotype (AAV6) . AAV6  
CC can be used to construct AAV viral vectors for use in gene therapy for a  
CC range of disorders: cystic fibrosis, cancer, AIDS, atherosclerosis,  
CC sickle cell anaemia, thalassaemia, blood clotting disorders and diabetes.  
CC The AAV viral vectors have increased transduction efficiency of a  
CC particular host cell as the AAV virion containing the AAV vector genome  
CC can be modified to express a capsid protein of an AAV serotype that  
CC transduces the selected host cell.

Sequence 4683 BP; 1114 A; 1363 C; 1277 G; 929 T; 0 other;

Query Match	80.3%	Score 114.8;	DB 22;	Length 4683;
Best Local Similarity	88.0%;	Pred. No. 4.1e-21;		
Matches 125; Conservative	0;	Mismatches 17;	Indels 0;	Gaps 0;

Oy 2 TACCCCTAGATGATGAGTTGCCACTCCCTCTGCGCGCTGCTCGCTCGGTGGGCGCT 61  
 |||||  
 Db 4542 TACCCCTAGATGATGAGTTGCCACTCCCTCTATGCGCGCTCGCTCGGTGGGCGCG 460

[illegible]

QY	122	CGCGCAGAGAGGAGTGGGCAA	143
Db	4662	CGCGCATAGAGCGAGTGGGCAA	4683

RESULT 7  
AAT63408/c  
ID AAT63408 standard; DNA; 145 BP.

AC AAT63408;

DT 18-DEC-1997 (first entry)

DE Inverted terminal repeat from adeno associated virus serotype 2.



KW Inverted terminal repeat; ITR; wild type; adeno associated virus;  
KW AAV; serotype 2; Rep binding site; integration system; treatment;  
KW acquired; inherited; disease; gene therapy; ss.  
XX  
OS Adeno associated virus.  
XX  
FH Key Location/Qualifiers  
FT repeat\_unit . 1..42  
FT /\*tag= a  
FT /rpt\_type= inverted  
FT /note= "region A"  
FT 85..125  
FT /\*tag= b  
FT /rpt\_type= inverted  
FT /note= "region A"  
FT 43..51  
FT /\*tag= c  
FT /rpt\_type= inverted  
FT /note= "region C"  
FT 53..61  
FT /\*tag= d  
FT /rpt\_type= inverted  
FT /note= "region C"  
FT 64..73  
FT /\*tag= e  
FT /rpt\_type= inverted  
FT /note= "region B"  
FT 75..84  
FT /\*tag= f  
FT /rpt\_type= inverted  
FT /note= "region B"  
FT 20..31  
FT /\*tag= g  
FT /label= Rep\_binding\_site  
FT 126..145  
FT /\*tag= h  
FT /note= "region D"  
XX  
XX WO9709442-A1.  
XX  
XX 13-MAR-1997.  
XX  
XX 06-SEP-1996; 96WO-US14312.  
XX  
XX 08-SEP-1995; 95US-0525866.  
XX  
XX (AVIG-) AVIGEN INC.  
XX  
XX Natsoulis G, Surosky RT;  
XX  
XX WPI; 1997-192917/17.  
XX  
XX  
PT Nucleic acid construct for integration into target site in recipient  
PT genome - comprises target sequence to bind AAV Rep and heterologous  
PT nucleotide sequence, for treatment of acquired or inherited disease  
XX  
XX  
PS Claim 3; Fig 1; 54pp; English.  
XX  
XX The present sequence is the inverted terminal repeat (ITR) from the  
XX wild type adeno associated virus (AAV) serotype 2 genome.  
XX A novel nucleic acid construct, comprises a targeting sequence  
XX capable of binding AAV Rep, e.g. the present sequence, and at least  
XX 1 heterologous nucleotide sequence arranged relative to the  
XX targeting sequence, so that it is integrated into a target site in  
XX a recipient genome. The integration system can be used to integrate  
XX a selected nucleotide sequence into a recipient cell genome, useful  
XX in the treatment of an acquired or inherited disease, e.g. long  
XX term gene therapy. The integration system has the site specific  
XX integration characteristics of AAV, is not limiting in regard to  
XX the size of the nucleotide sequence to be delivered and does not  
XX result in the integration of viral gene sequences into the  
XX recipient cell genome. Specifically it can be used to safely and  
XX efficiently deliver, and integrate large nucleotide sequences into

CC a suitable recipient cell genome without the risk of causing  
CC insertional mutagenesis due to random integration events.  
XX  
XX  
SQ Sequence 145 BP; 20 A; 52 C; 49 G; 24 T; 0 other;  
Query Match 75.1%; Score 107.4; DB 18; Length 145;  
Best Local Similarity 85.1%; Pred. No. 3e-19;  
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;  
OY 3 ACCCTAGTATGGAGTTGCCCACTCCTCTCTGCGCGCTCGCTGCTGGGGCCCTG 62  
DB 141 ACCCTAGTATGGAGTTGGCCACTCCTCTCTGCGCGCTCGCTGCTGAGGCCCG 82  
OY 63 CGACCAAGGTCGCGACGCGCAGAGCTCTGCTGCGCGGCCCAAGGAGCGAGC 122  
DB 81 GCGACCAAGGTCGCGCGCGCGCTTTGCCCGGCGGCTTCACTGAGCGAGCGAGC 22  
OY 123 GCGCAGAGGAGGTGGCAA 143  
DB 21 GCGCAGAGGAGGTGGCAA 1  
RESULT 8  
ABA02990 ID ABA02990 standard; DNA; 145 BP.  
XX  
XX ABA02990;  
XX AC  
XX 19-FEB-2002 (first entry)  
XX  
XX Adeno-associated virus 2 ITR SEQ ID NO 2.  
XX  
XX Cyrostatic; virucide; apoptosis inducer; p53; Saos-2 cell; infection;  
KW cancer; virus; HPV16; HPV18; adeno-associated virus 2; AAV-2; ss.  
XX  
XX Adeno associated virus 2.  
XX  
XX  
XX Key Location/Qualifiers  
FH misc\_structure 1..145  
FT /\*tag= a  
FT /label= "ITR"  
FT 72  
FT /\*tag= b  
FT /note= "unpaired base"  
FT 94  
FT /\*tag= c  
FT /note= "unpaired base"  
XX  
XX misc\_feature  
XX  
XX  
XX WO200180840-A2.  
XX  
XX  
XX 01-NOV-2001.  
XX  
XX 20-APR-2001; 2001WO-GB01795.  
XX  
XX 20-APR-2000; 2000GB-0009887.  
XX  
XX (BTGI-) BTG INT LTD.  
XX  
XX Raj K, Beard PM;  
XX  
XX WPI; 2002-041365/05.  
XX  
XX Single stranded and/or looped DNA for treating mutant p53 associated  
XX cancer or infection that inhibit cellular p53, having portion with an  
XX base, internally located with respect to any 3' and 5' ends of the DNA  
XX  
XX Example 11; Page 38; 51pp; English.  
XX  
XX The invention relates to single stranded and/or looped DNA having a  
XX portion with at least one base, internally located with respect to any 3'  
XX and 5' ends of the DNA, that is unbasepaired with another base in a form  
XX that is capable of being internalised within a target cell, for use in



FT /label= B\_repeat  
FT repeat\_unit 96..103  
FT /tag= g  
FT /rpt\_type= INVERTED  
FT repeat\_unit /label= B'\_repeat  
FT 104..145  
FT /tag= h  
FT /rpt\_type= INVERTED  
FT /label= A\_repeat  
FT /note= "Forms the stem of a T-shaped structure when  
FT base paired with A' repeat"  
FT repeat\_unit 146..165  
FT /tag= i  
FT /label= D'  
FT /note= "Additional D' sequence"  
FT 61..82  
FT /tag= j  
FT /note= "Part of T-shaped structure, formed by base  
FT pairing between the C' and C repeats"  
FT stem\_loop 84..104  
FT /tag= k  
FT /note= "Part of T-shaped structure, formed by base  
FT pairing between the B' and B repeats"  
FT misc\_structure 21..145  
FT /tag= l  
FT /label= T-shaped  
FT /note= "The naturally occurring 145 bp ITR sequence  
FT can form a T-shaped structure for DNA  
FT replication when single-stranded"  
FT repeat\_region 1..165  
FT /tag= m  
FT /label= Double-D  
FT /note= "This 165 bp sequence, resulting from  
FT addition of a second D repeat to the AAV  
FT ITR, has not been identified in any  
FT naturally occurring virus"  
PN WO9636364-A1.  
XX 21-NOV-1996.  
XX 14-MAY-1996; 96WO-US06786.  
XX 15-MAY-1995; 95US-0440738.  
PA (SAMU/) SAMULSKI R J.  
PA (XIAO/) XIAO X.  
PI Samulski RJ, Xiao X;  
XX WPI; 1997-042643/04.  
XX  
PT Double-D sequence directs adeno-associated virus integration into  
PT host genome - used in gene therapy, maintains full length coding  
PT sequence of therapeutic gene  
XX  
PS Claim 1; Fig 9; 55pp; English.  
XX  
CC The adeno associated virus (AAV) has a 145 bp inverted terminal  
CC repeat (ITR) located at each end of its genome. In addition to  
CC being able to base pair with each other, the ITRs can also  
CC individually fold back on themselves through the base pairing of A,  
CC A', B, B' and C, C' sequences to form a T-shaped structure for DNA  
CC replication (see features table). It has been found that viral  
CC mutants with deleted D sequences are unable to replicate their DNA.  
CC Addition of a second D repeat (D') to the naturally occurring ITR  
CC resulted in a sequence, designated double-D, which was sufficient  
CC to carry out the functions normally required of two wild-type ITRs  
CC during a lytic AAV viral infection, i.e. it is capable of directing  
CC replication and assembly into AAV, and/or the integration into the  
CC host genome, of recombinant DNA containing the nucleic acid molecule.  
CC Vectors and viral particles containing the double-D sequence are  
CC useful in gene therapy. Replication and integration into the host

CC Genome is completely effected through the double-D sequences, ensuring  
CC that the heterologous gene sequences remain intact.  
SQ Sequence 165 BP; 27 A; 56 C; 56 G; 26 T; 0 other;  
Query Match 75.1%; Score 107.4; DB 18; Length 165;  
Best Local Similarity 85.1%; Pred. No. 3e-19;  
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;  
QY 3 ACCCTAGTATGAGTGTGCCACTCCTCTCTGCGCGCTGCTGCTGCTGGGCGCTG 62  
DB 5 ACCCTAGTATGAGTGTGCCACTCCTCTCTGCGCGCTGCTGCTGCTGAGGCCGG 64  
QY 63 CGACCAAGGTCGCGACGCGGAGCTCTGCTGCGCGCGCGCCGACGAGCGAGCGAGC 122  
DB 65 GCGACCAAGGTCGCGCGACGCGCGGCTTGGCCGCGGCTGCTGAGCGAGCGAGC 124  
QY 123 GCGCAGAGGAGGAGTGGCAA 143  
DB 125 GCGCAGAGGAGGAGTGGCAA 145  
RESULT 11  
AAD03535  
ID AAD03535 standard; DNA; 165 BP.  
XX  
AC AAD03535;  
XX  
DT 19-JUN-2001 (first entry)  
XX  
DE Inverted terminal repeat double DD DNA sequence.  
XX  
KW Recombinant adeno-associated virus; rAAV; circular intermediate; ITR;  
KW inverted terminal repeat; haemostatic; antisickling; neuroprotective;  
KW antianaemic; nootropic; blood disorder; sickle cell anaemia;  
KW thalassaemia; neurological disorder; haemophilia; Alzheimer's disease;  
KW muscle disorder; Parkinson's disease; gene delivery; erythropoietin;  
KW epo; CFTR; cystic fibrosis transmembrane conductance receptor;  
KW tyrosine hydroxylase; glucocerebrosidase gene; Gaucher's disease;  
KW gene therapy; Double DD; ds.  
XX  
OS Unidentified.  
XX  
XX  
PN WO200125465-A1.  
XX  
PD 12-APR-2001.  
XX  
XX  
PF 06-OCT-2000; 2000WO-US27863.  
XX  
PR 07-OCT-1999; 99US-0158209.  
XX  
XX (IOWA ) UNIV IOWA RES FOUND.  
PA (ENGELHARDT J F.  
PA (DONG/) DONGSHENG D.  
PA (ZIYI/) ZIYING Y.  
PI Engelhardt JF, Dongsheng D, Ziyang Y;  
XX WPI; 2001-266321/27.  
XX  
XX Composition for transferring recombinant DNAs and to express a  
PT polypeptide in a host cell, comprises two recombinant adeno-associated  
PT viruses -  
XX  
PS Disclosure; Page 139; 144pp; English.  
XX  
CC The patent discloses a composition comprising at least two recombinant  
CC adeno-associated viruses (rAAVs). Each rAAV contains a recombinant DNA  
CC molecule comprising three DNA segments linked together. The first and  
CC the third DNA segments comprise 5' and 3' inverted terminal repeats  
CC (ITRs) respectively from the circular intermediate of AAV. The second  
CC DNA segment in each virus is different and does not comprise AAV  
CC sequence. It preferably comprises sequences encoding a therapeutically





CC vector with therapeutic gene and delivery vehicle or containing two  
CC vectors expressing a full length polypeptide coordinately, are useful  
CC for manufacturing a medicament for treating pathological conditions or  
CC symptoms in a mammal. The vector is useful for therapeutic or  
CC prophylactic treatments of blood disorders (e.g. sickle cell anemia,  
CC thalassemias, hemophilias and Fanconi's anemia), neurological disorders  
CC (e.g. Alzheimer's disease, Parkinson's disease) and muscle disorders.  
XX  
SQ Sequence 272 BP; 58 A; 78 C; 78 G; 58 T; 0 other;  
  
Query Match 75.1%; Score 107.4; DB 21; Length 272;  
Best Local Similarity 85.1%; Pred. No. 3.1e-19;  
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;  
  
QY 3 ACCCCTAGTATGAGTTGCCCACTCCTCTCTGCGGCTGCTGCTGCTGCTGCTGCTG 62  
DB 53 ACCCTAGTATGAGTTGGCCACTCCCTCTCTGCGGCTGCTGCTGCTGCTGCTGCTG 112  
QY 63 CGACCAAGGTCGCGACGCGCAGAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 122  
DB 113 GCGACCAAGGTCGCGCGACGCGGCTTTGCTGCGCGCTGCTGCTGCTGCTGCTG 172  
QY 123 GCGCAGAGAGGAGTGGGCAA 143  
DB 173 GCGCAGAGAGGAGTGGCCAA 193  
  
RESULT 14  
AAD03534  
ID AAD03534 standard; DNA; 272 BP.  
XX  
AC AAD03534;  
XX  
DT 19-JUN-2001 (first entry)  
XX  
DE AAV circular intermediate DNA segment comprising ITRs from clone p1202.  
XX  
KM Recombinant adeno-associated virus; rAAV; circular intermediate; ITR;  
KM inverted terminal repeat; haemostatic; antistickling; neuroprotective;  
KM antihaemic; neurotropic; blood disorder; sickle cell anaemia;  
KM thalassemia; neurological disorder; haemophilia; Alzheimer's disease;  
KM muscle disorder; Parkinson's disease; gene delivery; erythropoietin;  
KM epo; CFTR; cystic fibrosis transmembrane conductance receptor;  
KM tyrosine hydroxylase; glucocerebrosidase gene; Gaucher's disease;  
KM gene therapy; clone p1202; ds.  
XX  
OS Adeno associated virus.  
XX  
PN WO200125465-A1.  
XX  
PD 12-APR-2001.  
XX  
PF 06-OCT-2000; 2000WO-US27863.  
XX  
PR 07-OCT-1999; 99US-0158209.  
XX  
PA (IOWA ) UNIV IOWA RES FOUND.  
PA (ENGE/) ENGELHARDT J F.  
PA (DONG/) DONGSHENG D.  
PA (ZIYI/) ZIYING Y.  
XX  
PI Engelhardt JF, Dongsheg D, Ziyng Y;  
XX  
DR WPI; 2001-266321/27.  
XX  
XX Composition for transferring recombinant DNAs and to express a  
PT polypeptide in a host cell, comprises two recombinant adeno-associated  
PT viruses -  
XX  
PS Disclosure; Fig 10C; 144pp; English.  
XX  
CC The patent discloses a composition comprising at least two recombinant  
CC adeno-associated viruses (rAAVs). Each rAAV contains a recombinant DNA

CC molecule comprising three DNA segments linked together. The first and  
CC the third DNA segments comprise 5' and 3' inverted terminal repeats  
CC (ITRs) respectively from the circular intermediate of AAV. The second  
CC DNA segment in each virus is different and does not comprise AAV  
CC sequence. It preferably comprises sequences encoding a therapeutically  
CC effective polypeptide such as the cystic fibrosis transmembrane  
CC conductance receptor gene (CFTR), the erythropoietin (epo) gene, the  
CC tyrosine hydroxylase gene (Parkinson's syndrome) or the glucocere-  
CC brosidase gene (Gaucher's disease). The circular intermediate of AAV  
CC imparts increased episomal stability and persistence of the vector in  
CC the host cell. Compositions comprising rAAV sequences are useful for  
CC transferring recombinant DNAs to a host cell and express a polypeptide  
CC in a host cell. The recombinant vector is useful in medical therapy,  
CC which includes treatment or prophylaxis of blood disorders (e.g. sickle  
CC cell anaemia, thalassemia, haemophilia), neurological disorders, such  
CC as Alzheimer's disease, Parkinson's disease, muscle disorders involving  
CC skeletal, cardiac or smooth muscle. AAV vector is used as a delivery  
CC vehicle for gene therapy.  
CC The present sequence is circular intermediate DNA segment comprising  
CC ITRs from Adeno associated virus (AAV) clone p1202.  
XX  
SQ Sequence 272 BP; 58 A; 78 C; 78 G; 58 T; 0 other;  
  
Query Match 75.1%; Score 107.4; DB 22; Length 272;  
Best Local Similarity 85.1%; Pred. No. 3.1e-19;  
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;  
  
QY 3 ACCCCTAGTATGAGTTGCCCACTCCTCTCTGCGGCTGCTGCTGCTGCTGCTGCTG 62  
DB 53 ACCCCTAGTATGAGTTGGCCACTCCTCTCTGCGGCTGCTGCTGCTGCTGCTGCTG 112  
QY 63 CGACCAAGGTCGCGACGCGCAGAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 122  
DB 113 GCGACCAAGGTCGCGCGACGCGGCTTTGCTGCGCGCTGCTGCTGCTGCTGCTG 172  
QY 123 GCGCAGAGAGGAGTGGGCAA 143  
DB 173 GCGCAGAGAGGAGTGGCCAA 193  
  
RESULT 15  
AAH41481  
ID AAH41481 standard; DNA; 4675 BP.  
XX  
AC AAH41481;  
XX  
DT 23-AUG-2001 (first entry)  
XX  
DE Adeno-associated virus (AAV) plasmid pAV1 Avaii nucleotide sequence.  
XX  
KM Adeno-associated virus; AAV; gene transfer; adenovirus; chromosome 19;  
KM medicine; cell engineering; genetic engineering; treatment;  
KM embryological engineering; ds.  
XX  
OS Adeno-associated virus.  
XX  
PN WO200132899-A1.  
XX  
PD 10-MAY-2001.  
XX  
PF 23-OCT-2000; 2000WO-JP07373.  
XX  
PR 29-OCT-1999; 99JP-0308839;  
XX  
PA (TAKI ) TAKARA SHUZO CO LTD.  
XX  
PI Ueno T, Matsumura H, Tanaka K, Iwasaki T, Ueno M, Fujinaga K;  
PI Asada K, Kato I;  
XX  
DR WPI; 2001-316450/33.  
XX  
XX Gene transfer method for highly efficient introduction of foreign genes  
PT to cells in humans particularly by integration specifically into AAVS1

PT site of chromosome 19 for transformation, applicable e.g. in disease  
PT treatment -

PS Example 1; Page 28-31; 38pp; Japanese.

CC The present invention describes a gene transfer method for highly  
CC efficient introduction of foreign genes to cells in humans particularly  
CC by integration specifically into AAVS1 site of chromosome 19 for  
CC transformation, applicable e.g. in disease treatment. The method  
CC comprises transferring into the cells (with the use of an adenovirus  
CC vector) a nucleic acid which has a sequence provided with adeno-  
CC associated virus (AAV)-originated inverted terminal repeats (ITRs) in  
CC both sides of the target foreign gene to be transferred, a second nucleic  
CC acid which has an AAV-originated rep gene and a promoter for expressing  
CC this gene and carries a stuffer sequence inserted into it sandwiched in  
CC two recombinase-recognition sequences and located between the rep gene  
CC and promoter and expressing the Rep protein under the action of the  
CC recombinase in the cells obtained previously to integrate the target  
CC foreign gene into the chromosomal DNA. The method is for transferring a  
CC foreign gene into cells particularly in human, especially by integration  
CC specifically into AAVS1 site of chromosome 19 for transformation, which  
CC is applicable in medicine, as well as cell, genetic and embryological  
CC engineering e.g. in disease treatment. The method is highly efficient,  
CC with use of adenovirus vectors, nucleic acids and other sequences  
CC including rep genes and promoters, by expressing rep proteins to  
CC integrate target/foreign gene into chromosomal DNA through action of  
CC recombinase. The present sequence represents an AAV plasmid pAV1 Avail  
CC nucleotide sequence which is used in an example from the present  
CC invention.

XX  
XX Sequence 4675 BP; 1198 A; 1262 C; 1251 G; 964 T; 0 other;

Query Match 75.1%; Score 107.4; DB 22; Length 4675;  
Best Local Similarity 85.1%; Pred. No. 3.5e-19;  
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 3 ACCCTAGTATGAGTGGCCACTCCTCTCTGCGGCTCGCTGCTGGTGGGCTG 62  
DB 4535 ACCCTAGTATGAGTGGCCACTCCTCTCTGCGGCTCGCTGCTGAGGCTCGG 4594

QY 63 CGACCAAGGTCCGACGCGCAGAGCTCTGCTGCGGCGCCCAACGAGCGAGC 122  
DB 4595 GCGACCAAGGTCCGACGCGCAGAGCTCTGCTGCGGCGCCCAACGAGCGAGC 4654

QY 123 GCGCAGAGGAGTGGCAA 143  
DB 4655 GCGCAGAGGAGTGGCAA 4675

Search completed: July 20, 2003, 09:39:48  
Job time : 33.1207 secs







DB 81 GCGACCAAGGTCGCCCCGACGCGGCTTTGCCCCGGGCTCAGTGAGCGAGCGAGC 22  
QY 123 GCGCAGAGAGGAGTGGGCAA 143  
DB 21 GCGCAGAGAGGAGTGGCCAA 1

## RESULT 2

US-07-989-841A-1

; Sequence 1, Application US/07989841A  
; Patent No. 5478745  
; GENERAL INFORMATION:  
; APPLICANT: Samulski, R. J.  
; APPLICANT: Xiao, X.  
; TITLE OF INVENTION: Recombinant Viral Vector System  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/989,841A  
; FILING DATE: On even date herewith  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 6636-013  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 790-9090  
; TELEFAX: (212) 869-8864/9741  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 165 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: unknown  
; MOLECULE TYPE: DNA (genomic)  
; US-07-989-841A-1

Query Match 75.1%; Score 107.4; DB 1; Length 165;  
Best Local Similarity 85.1%; Pred. No. 1.8e-21;  
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 3 ACCCCTAGTATGAGTGTGCCCACTCCCTCTCTGCGCGCTCGCTCGGTGGGGCCTG 62  
DB 5 ACCCTAGTATGAGTGTGCCCACTCCCTCTCTGCGCGCTCGCTCGTCACTGAGGCCG 64  
QY 63 CGACCAAGGTCGCGACGCGCAGAGCTCTGCTCGCGCGCGCGCGCGCGCGCGCGAGC 122  
DB 65 GCGACCAAGGTCGCGCGACGCGCGCGCTTTGCGCGCGCGCGCGCTCAGTGAGCGAGC 124  
QY 123 GCGCAGAGAGGAGTGGGCAA 143  
DB 125 GCGCAGAGAGGAGTGGCCAA 145

## RESULT 3

US-08-440-738A-1

; Sequence 1, Application US/08440738A  
; Patent No. 5869305  
; GENERAL INFORMATION:  
; APPLICANT: Samulski, R. J.

; APPLICANT: Xiao, X.  
; TITLE OF INVENTION: Recombinant Viral Vector System  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/440,738A  
; FILING DATE: May 15, 1995.  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 6636-022  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 790-9090  
; TELEFAX: (212) 869-8864/9741  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 165 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: unknown  
; MOLECULE TYPE: DNA (genomic)  
; US-08-440-738A-1

Query Match 75.1%; Score 107.4; DB 2; Length 165;  
Best Local Similarity 85.1%; Pred. No. 1.8e-21;  
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 3 ACCCCTAGTATGAGTGTGCCCACTCCCTCTCTGCGCGCTCGCTCGGTGGGGCCTG 62  
DB 5 ACCCCTAGTATGAGTGTGCCCACTCCCTCTCTGCGCGCTCGCTCGTCACTGAGGCCG 64  
QY 63 CGACCAAGGTCGCGACGCGCAGAGCTCTGCTCGCGCGCGCGCGCGCGCGCGAGC 122  
DB 65 GCGACCAAGGTCGCGCGACGCGCGCGCTTTGCGCGCGCGCGCGCTCAGTGAGCGAGC 124  
QY 123 GCGCAGAGAGGAGTGGGCAA 143  
DB 125 GCGCAGAGAGGAGTGGCCAA 145

## RESULT 4

US-08-471-914-1

; Sequence 1, Application US/08471914A  
; Patent No. 6057152  
; GENERAL INFORMATION:  
; APPLICANT: Samulski, R.  
; APPLICANT: Xiao, X.  
; TITLE OF INVENTION: RECOMBINANT VIRAL VECTOR SYSTEM  
; FILE REFERENCE: 6636-027  
; CURRENT APPLICATION NUMBER: US/08/471,914A  
; EARLIER APPLICATION NUMBER: 08/440,738  
; EARLIER FILING DATE: 1995-05-15  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 165  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:



OTHER INFORMATION: Description of Artificial Sequence: double-D  
OTHER INFORMATION: sequence  
US-08-471-914-1

Query Match 75.1%; Score 107.4; DB 3; Length 165;  
Best Local Similarity 85.1%; Pred. No. 1.8e-21;  
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 3 ACCCTAGTATGAGTGTCCCACTCCCTCTCTGCGGCTCGCTCGCTCGGCGGCTG 62  
|||||  
DB 5 ACCCTAGTATGAGTGTGCGCACTCCCTCTCTGCGGCTCGCTCGCTCACTGAGGCGG 64  
QY 63 CGACCAAGGTCCGACAGCGGAGAGCTCTGCTGCGGCGGCGGCGGCGGCGGAGC 122  
|||||  
DB 65 GCGACCAAGGTCCGCGCGCGGCTTGTGCGGCGGCGGCGGCGGCTCACTGAGCGAGC 124  
QY 123 GCGCAGAGGAGGTGGCAA 143  
|||||  
DB 125 GCGCAGAGGAGGTGGCAA 145

## RESULT 5

US-09-276-625-7  
Sequence 7, Application US/09276625  
Patent No. 6436392  
GENERAL INFORMATION:  
APPLICANT: Engelhardt, John F.  
APPLICANT: Duan, Dongsheng  
TITLE OF INVENTION: Adeno-associated virus vectors  
FILE REFERENCE: 875.007US1  
CURRENT APPLICATION NUMBER: US/09/276,625  
CURRENT FILING DATE: 1999-03-25  
PRIOR APPLICATION NUMBER: US 60/086,166  
PRIOR FILING DATE: 1998-05-20  
NUMBER OF SEQ ID NOS: 13  
SOFTWARE: FastSeq for windows Version 4.0  
SEQ ID NO 7  
LENGTH: 165  
TYPE: DNA  
ORGANISM: Unknown  
FEATURE:  
OTHER INFORMATION: SEQ ID NO:1 of U.S. Patent No. 6436392 5,478,745  
US-09-276-625-7

Query Match 75.1%; Score 107.4; DB 4; Length 165;  
Best Local Similarity 85.1%; Pred. No. 1.8e-21;  
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 3 ACCCTAGTATGAGTGTCCCACTCCCTCTCTGCGGCTCGCTCGCTCGGCGGCTG 62  
|||||  
DB 5 ACCCTAGTATGAGTGTGCGCACTCCCTCTCTGCGGCTCGCTCGCTCACTGAGGCGG 64  
QY 63 CGACCAAGGTCCGACAGCGGAGAGCTCTGCTGCGGCGGCGGCGGCGGCGGAGC 122  
|||||  
DB 65 GCGACCAAGGTCCGCGCGGCTTGTGCGGCGGCGGCGGCGGCTCACTGAGCGAGC 124  
QY 123 GCGCAGAGGAGGTGGCAA 143  
|||||  
DB 125 GCGCAGAGGAGGTGGCAA 145

## RESULT 6

US-08-702-573-3  
Sequence 3, Application US/08702573  
Patent No. 6033885  
GENERAL INFORMATION:  
APPLICANT: LATTA, Martine  
APPLICANT: DENEFL, Patrice  
APPLICANT: VIGNE, Emmanuelle  
APPLICANT: PERRICAUDET, Michel  
TITLE OF INVENTION: INTEGRATIVE RECOMBINANT ADENOVIRUSES,  
TITLE OF INVENTION: PREPARATION THEREOF AND THERAPEUTICAL USES THEREOF  
NUMBER OF SEQUENCES: 13

## CORRESPONDENCE ADDRESS:

ADDRESSEE: Rhone-Poulenc Rorer Inc.  
STREET: 500 Arcola Rd. 3C43  
CITY: Collegeville  
STATE: PA  
COUNTRY: USA  
ZIP: 19426

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/702,573  
FILING DATE:

## CLASSIFICATION: 424

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 94/02445  
FILING DATE: 03-MAR-1994

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: WO PCT/FR95/00233  
FILING DATE: 28-FEB-1995

## ATTORNEY/AGENT INFORMATION:

NAME: Smith Ph.D., Julie K.  
REGISTRATION NUMBER: 38,619

REFERENCE/DOCKET NUMBER: ST94011-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610)454-3839

TELEFAX: (610)454-3808  
INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:  
LENGTH: 192 base pairs

TYPE: nucleic acid  
STRANDEDNESS: double

TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid

FEATURE:  
NAME/KEY: misc feature

LOCATION: 1..192  
OTHER INFORMATION: /note= "Right ITR Sequence in

Patent No. 6033885  
OTHER INFORMATION: PXL2384"

US-08-702-573-3

Query Match 75.1%; Score 107.4; DB 3; Length 192;  
Best Local Similarity 85.1%; Pred. No. 1.8e-21;  
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 3 ACCCTAGTATGAGTGTCCCACTCCCTCTCTGCGGCTCGCTCGCTCGGCGGCTG 62  
|||||  
DB 52 ACCCTAGTATGAGTGTGCGCACTCCCTCTCTGCGGCTCGCTCGCTCACTGAGGCGG 111  
QY 63 CGACCAAGGTCCGACAGCGGAGAGCTCTGCTGCGGCGGCGGCGGCGGCGGAGC 122  
|||||  
DB 112 GCGACCAAGGTCCGCGCGGCTTGTGCGGCGGCGGCGGCGGCTCACTGAGCGAGC 171  
QY 123 GCGCAGAGGAGGTGGCAA 143  
|||||  
DB 172 GCGCAGAGGAGGTGGCAA 192

## RESULT 7

US-09-276-625-6  
Sequence 6, Application US/09276625  
Patent No. 6436392  
GENERAL INFORMATION:  
APPLICANT: Engelhardt, John F.  
APPLICANT: Duan, Dongsheng  
TITLE OF INVENTION: Adeno-associated virus vectors  
FILE REFERENCE: 875.007US1  
CURRENT APPLICATION NUMBER: US/09/276,625  
CURRENT FILING DATE: 1999-03-25  
PRIOR APPLICATION NUMBER: US 60/086,166



```
RESULT 10
; US-08-709-609-1
; Sequence 1, Application US/08709609
; Patent No. 5858775
; GENERAL INFORMATION:
; APPLICANT: Johnson, Philip R.
; TITLE OF INVENTION: Adeno-Associated Virus Materials and
; TITLE OF INVENTION: Methods
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 S. Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/709,609
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5858775and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 31975
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4680 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-709-609-1

Query Match      75.1%; Score 107.4; DB 2; Length 4680;
Best Local Similarity 85.1%; Pred. No. 2.5e-21;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY      3 ACCCTAGTATGAGTGTGCCCACTCCCTCTCTGCGGCTGCTGCTGGGCTG 62
DB      4540 ACCCTAGTATGAGTGTGCCCACTCCCTCTCTGCGGCTGCTGCTGAGGCCG 4599

QY      63 CGACCAAGGTCCGACAGCGGAGAGCTCTGCTGCGGCGCCCGACGAGCGAGC 122
DB      4600 GCGACCAAGGTCCGACGCGCGGCTTGTGCGGCGGCGCTCACTGAGCGAGC 4659

QY      123 GCGCAGAGAGGAGTGGCAA 143
DB      4660 GCGCAGAGAGGAGTGGCAA 4680

RESULT 11
; PCT-US95-07178-1
; Sequence 1, Application PC/TUS9507178
; GENERAL INFORMATION:
; APPLICANT: Johnson, Philip R.
; TITLE OF INVENTION: Adeno-Associated Virus Materials and
; TITLE OF INVENTION: Methods
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 S. Wacker Drive
; CITY: Chicago
; STATE: Illinois
```

```
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/07178
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Noland, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 31975
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4680 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; PCT-US95-07178-1
```

```
Query Match      75.1%; Score 107.4; DB 5; Length 4680;
Best Local Similarity 85.1%; Pred. No. 2.5e-21;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
```

```
QY      3 ACCCTAGTATGAGTGTGCCCACTCCCTCTCTGCGGCTGCTGCTGGGCTG 62
DB      4540 ACCCTAGTATGAGTGTGCCCACTCCCTCTCTGCGGCTGCTGCTGAGGCCG 4599

QY      63 CGACCAAGGTCCGACAGCGGAGAGCTCTGCTGCGGCGCCCGACGAGCGAGC 122
DB      4600 GCGACCAAGGTCCGACGCGCGGCTTGTGCGGCGGCGCTCACTGAGCGAGC 4659

QY      123 GCGCAGAGAGGAGTGGCAA 143
DB      4660 GCGCAGAGAGGAGTGGCAA 4680
```

```
RESULT 12
; US-09-770-315-2
; Sequence 2, Application US/09770315
; Patent No. 6429001
; GENERAL INFORMATION:
; APPLICANT: Chiron Corporation
; TITLE OF INVENTION: Recombinant AAV Packaging Systems
; FILE REFERENCE: 20263-501
; CURRENT APPLICATION NUMBER: US/09/770,315
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: US 60/178,536
; PRIOR FILING DATE: 2000-01-26
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 8698
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: recombinant DNA
; US-09-770-315-2
```

```
Query Match      75.1%; Score 107.4; DB 4; Length 8698;
Best Local Similarity 85.1%; Pred. No. 2.7e-21;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY      3 ACCCTAGTATGAGTGTGCCCACTCCCTCTCTGCGGCTGCTGCTGGGCTG 62
DB      4540 ACCCTAGTATGAGTGTGCCCACTCCCTCTCTGCGGCTGCTGCTGAGGCCG 4599
```



Db 4539 ACCCTAGTATGAGTGGCCACTCCCTCTCTGCGGCTGCTGCTCACTGAGGCCG 4598  
QY 63 CGACCAAGTCCGCGACGCGAGAGCTCTGCTGCGGCCCCACCGAGCGAGC 122  
Db 4599 GCGACCAAGTCCCGCGACGCGGCTTGTGCGCGCGGCTCACTGAGCGAGCGAGC 4658  
QY 123 GCGCAGAGAGGAGTGGGCAA 143  
Db 4659 GCGCAGAGAGGAGTGGCCAA 4679

## RESULT 13

US-09-276-625-4  
; Sequence 4, Application US/09276625  
; Patent No. 6436392  
; GENERAL INFORMATION:  
; APPLICANT: Engelhardt, John F.  
; APPLICANT: Duan, Dongsheng  
; TITLE OF INVENTION: Adeno-associated virus vectors  
; FILE REFERENCE: 875.007US1  
; CURRENT APPLICATION NUMBER: US/09/276,625  
; PRIOR FILING DATE: 1999-03-25  
; PRIOR APPLICATION NUMBER: US 60/086,166  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 272  
; TYPE: DNA  
; ORGANISM: AAV circular intermediate, clone p81  
US-09-276-625-4

Query Match 74.0%; Score 105.8; DB 4; Length 272;  
Best Local Similarity 84.4%; Pred. No. 5.2e-21;  
Matches 119; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 3 ACCCTAGTATGAGTGGCCACTCCCTCTCTGCGGCTGCTGCTCGGTGGGCTG 62  
Db 53 ACCCTAGTATGAGTGGCCACTCCCTCTCTGCGGCTGCTGCTCACTGAGGCCG 112  
QY 63 CGACCAAGTCCGCGACGCGAGAGCTCTGCTGCGGCCCCACCGAGCGAGC 122  
Db 113 GCGCACAAGTCCCGCGACGCGGCTTGTGCGCGGCGGCTCACTGAGCGAGC 172  
QY 123 GCGCAGAGAGGAGTGGGCAA 143  
Db 173 GCGCAGAGAGGAGTGGCCAA 193

## RESULT 14

US-09-394-110A-1/c  
; Sequence 1, Application US/09394110A  
; Patent No. 6451594  
; GENERAL INFORMATION:  
; APPLICANT: Chien, Kenneth  
; APPLICANT: Wang, Yibin  
; APPLICANT: Evans, Sylvia  
; TITLE OF INVENTION: No. 6451594e1 Recombinant Adenovirus for Tissue Specific Expression  
; FILE REFERENCE: 6627-PA8045  
; CURRENT APPLICATION NUMBER: US/09/394,110A  
; CURRENT FILING DATE: 1999-09-10  
; NUMBER OF SEQ ID NOS: 3  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 1  
; LENGTH: 174  
; TYPE: DNA  
; ORGANISM: adeno-associated virus 2  
US-09-394-110A-1

Query Match 73.7%; Score 105.4; DB 4; Length 174;  
Best Local Similarity 84.9%; Pred. No. 6.5e-21;  
Matches 118; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 3 ACCCTAGTATGAGTGGCCACTCCCTCTCTGCGGCTGCTGCTCGGTGGGCTG 62  
Db 139 ACCCTAGTATGAGTGGCCACTCCCTCTCTGCGGCTGCTGCTCACTGAGGCCG 80  
QY 63 CGACCAAGTCCGCGACGCGAGAGCTCTGCTGCGGCCCCACCGAGCGAGC 122  
Db 79 GCGACCAAGTCCCGCGACGCGGCTTGTGCGCGGCGGCTCACTGAGCGAGCGAGC 20  
QY 123 GCGCAGAGAGGAGTGGG 141  
Db 19 GCGCAGAGAGGAGTGGCC 1

## RESULT 15

US-09-276-625-6/c  
; Sequence 6, Application US/09276625  
; Patent No. 6436392  
; GENERAL INFORMATION:  
; APPLICANT: Engelhardt, John F.  
; APPLICANT: Duan, Dongsheng  
; TITLE OF INVENTION: Adeno-associated virus vectors  
; FILE REFERENCE: 875.007US1  
; CURRENT APPLICATION NUMBER: US/09/276,625  
; PRIOR FILING DATE: 1999-03-25  
; PRIOR APPLICATION NUMBER: US 60/086,166  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6  
; LENGTH: 272  
; TYPE: DNA  
; ORGANISM: AAV circular intermediate, clone p1202  
US-09-276-625-6

Query Match 71.7%; Score 102.6; DB 4; Length 272;  
Best Local Similarity 83.0%; Pred. No. 4e-20;  
Matches 117; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 3 ACCCTAGTATGAGTGGCCACTCCCTCTCTGCGGCTGCTGCTCGGTGGGCTG 62  
Db 209 ACCCTAGTATGAGTGGCCACTCCCTCTCTGCGGCTGCTGCTCACTGAGGCCG 150  
QY 63 CGACCAAGTCCGCGACGCGAGAGCTCTGCTGCGGCCCCACCGAGCGAGC 122  
Db 149 GCGACCAAGTCCCGCGAGAGCTTGTGCGCGGCGGCTCACTGAGCGAGCGAGC 90  
QY 123 GCGCAGAGAGGAGTGGGCAA 143  
Db 89 GCGCAGAGAGGAGTGGCCAA 69

Search completed: July 20, 2003, 13:23:49  
Job time : 7.68705 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2003, 13:19:56 ; Search time 30.7204 Seconds  
(without alignments)  
9603.029 Million cell updates/sec

Title: US-09-807-802A-1\_COPY\_4576\_4718

Perfect score: 143  
Sequence: 1 ttaccctagtgtgaggtt.....cgagagagggagtgagca 143

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1439767 seqs, 1031500376 residues

Total number of hits satisfying chosen parameters: 2879534

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA:\*

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq1:\*  
11: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq2:\*  
12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq3:\*  
13: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*  
15: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
16: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
17: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	107.4	75.1	145	US-10-240-198-2	Sequence 2, Appli
2	107.4	75.1	165	US-10-054-665-7	Sequence 7, Appli
3	107.4	75.1	207	US-10-023-208-58	Sequence 58, Appli
4	107.4	75.1	272	US-10-054-665-6	Sequence 6, Appli
5	107.4	75.1	4675	US-09-782-378A-1	Sequence 1, Appli
6	107.4	75.1	4675	US-09-782-378A-2	Sequence 2, Appli
7	107.4	75.1	4675	US-10-240-198-1	Sequence 1, Appli
8	107.4	75.1	4679	US-09-804-898-1	Sequence 1, Appli
9	107.4	75.1	4679	US-09-945-681-10	Sequence 10, Appli
10	107.4	75.1	4679	US-10-038-972A-12	Sequence 12, Appli
11	107.4	75.1	4680	US-10-077-294-1	Sequence 1, Appli
12	107.4	75.1	4680	US-10-163-886-1	Sequence 1, Appli
13	107.4	75.1	4680	US-10-263-127-1	Sequence 1, Appli
14	105.8	74.0	272	US-10-054-665-4	Sequence 4, Appli
15	102.6	71.7	272	US-10-054-665-6	Sequence 6, Appli
16	99.4	69.5	145	US-09-782-378A-6	Sequence 6, Appli

C 17	99.4	69.5	146	15	US-10-135-984-8	Sequence 8, Appli
C 18	99.4	69.5	165	11	US-09-782-378A-8	Sequence 8, Appli
C 19	99.4	69.5	165	15	US-10-054-665-7	Sequence 7, Appli
C 20	99.4	69.5	207	15	US-10-023-208-58	Sequence 58, Appli
C 21	99.4	69.5	4675	11	US-09-782-378A-1	Sequence 1, Appli
C 22	99.4	69.5	4675	11	US-09-782-378A-2	Sequence 2, Appli
C 23	99.4	69.5	4675	15	US-10-240-198-1	Sequence 1, Appli
C 24	99.4	69.5	4679	10	US-09-804-898-1	Sequence 1, Appli
C 25	99.4	69.5	4679	10	US-09-945-681-10	Sequence 10, Appli
C 26	99.4	69.5	4679	15	US-10-038-972A-12	Sequence 12, Appli
C 27	99.4	69.5	4680	15	US-10-077-294-1	Sequence 1, Appli
C 28	99.4	69.5	4680	15	US-10-163-886-1	Sequence 1, Appli
C 29	99.4	69.5	4680	15	US-10-263-127-1	Sequence 1, Appli
C 30	99.4	69.5	5932	15	US-10-267-117-4	Sequence 4, Appli
C 31	99.4	69.5	5932	15	US-10-267-117-4	Sequence 4, Appli
C 32	99.4	69.5	5932	15	US-10-340-112-4	Sequence 4, Appli
C 33	99.4	69.5	5932	15	US-10-267-117-8	Sequence 4, Appli
C 34	99.4	69.5	6142	15	US-10-267-117-8	Sequence 8, Appli
C 35	99.4	69.5	6142	15	US-10-267-117-8	Sequence 8, Appli
C 36	99.4	69.5	6142	15	US-10-340-112-8	Sequence 8, Appli
C 37	99.4	69.5	6142	15	US-10-340-112-8	Sequence 8, Appli
C 38	99.4	69.5	6565	15	US-10-267-117-1	Sequence 1, Appli
C 39	99.4	69.5	6565	15	US-10-267-117-1	Sequence 1, Appli
C 40	99.4	69.5	6565	15	US-10-340-112-1	Sequence 1, Appli
C 41	99.4	69.5	6565	15	US-10-340-112-1	Sequence 1, Appli
C 42	99.4	69.5	6714	15	US-10-267-117-6	Sequence 6, Appli
C 43	99.4	69.5	6714	15	US-10-267-117-6	Sequence 6, Appli
C 44	99.4	69.5	6714	15	US-10-340-112-6	Sequence 6, Appli
C 45	99.4	69.5	6714	15	US-10-340-112-6	Sequence 6, Appli

## ALIGNMENTS

RESULT 1  
US-10-240-198-2  
; Sequence 2, Application US/10240198  
; Publication No. US20030100115A1  
; GENERAL INFORMATION:  
; APPLICANT: BTG International Ltd  
; APPLICANT: BEARD DR, PETER  
; APPLICANT: RAJ DR, KENNETH  
; TITLE OF INVENTION: CYTOTOXIC AGENTS  
; FILE REFERENCE: 142184WO  
; CURRENT APPLICATION NUMBER: US/10/240,198  
; CURRENT FILING DATE: 2002-09-30  
; PRIOR APPLICATION NUMBER: 0009887.1  
; PRIOR FILING DATE: 2000-04-20  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 145  
; TYPE: DNA  
; ORGANISM: adeno-associated virus 2  
; FEATURE:  
; NAME/KEY: misc\_structure  
; LOCATION: (1)..(145)  
; OTHER INFORMATION: ITR  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (72)  
; OTHER INFORMATION: Unpaired base  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (94)  
; OTHER INFORMATION: Unpaired base  
US-10-240-198-2

Query Match 75.1%; Score 107.4; DB 15; Length 145;  
Best Local Similarity 85.1%; Pred. No. 2.6e-21;  
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;  
Qy 3 ACCCCTAGTGTGAGTTGCCCACTCCCTCTCTGCGGCTCGCTCGGTGGGCTG 62

```
Db 5 ACCCTAGTATGATGAGTTGGCCACTCCCTCTCTGCGGCTCGCTCGCTCACTGAGCCGG 64
QY 63 CGGACCAAGGTCGCGACGCGCAGAGCTCTGCTCTGCGGCGCCCGCCACCGCAGCGAGC 122
Db 65 GCGACCAAGGTCGCGCGACGCGCGGCTTTGCGCGCGGCGCTCACTGAGCGAGCGAGC 124
QY 123 GCGCAGAGAGGAGTGGGCAA 143
Db 125 GCGCAGAGAGGAGTGGCCAA 145
```

## RESULT 2

```
US-10-054-665-7
; Sequence 7, Application US/10054665
; Publication No. US20020197237A1
; GENERAL INFORMATION:
; APPLICANT: Engelhardt, John F.
; APPLICANT: Duan, Dongsheng
; TITLE OF INVENTION: Adeno-associated virus vectors
; FILE REFERENCE: 875.007US2
; CURRENT APPLICATION NUMBER: US/10/054,665
; PRIOR FILING DATE: 2002-06-13
; PRIOR APPLICATION NUMBER: US 09/276,625
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/086,166
; PRIOR FILING DATE: 1998-05-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 165
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: SEQ ID NO:1 of U.S. Patent No. US20020197237A1 5,478,745
US-10-054-665-7
```

```
Query Match 75.1%; Score 107.4; DB 15; Length 165;
Best Local Similarity 85.1%; Pred. No. 2.5e-21;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
```

```
QY 3 ACCCTAGTATGATGAGTTGCCACTCCCTCTCTGCGGCTCGCTCGCTCGGTGGGCGCTG 62
Db 5 ACCCTAGTATGATGAGTTGCCACTCCCTCTCTGCGGCTCGCTCGCTCACTGAGCGCGG 64
QY 63 CGGACCAAGGTCGCGACGCGCAGAGCTCTGCTCTGCGGCGCCCGCCACCGCAGCGAGC 122
Db 65 GCGACCAAGGTCGCGCGACGCGCGGCTTTGCGCGCGGCGCTCACTGAGCGAGCGAGC 124
QY 123 GCGCAGAGAGGAGTGGGCAA 143
Db 125 GCGCAGAGAGGAGTGGCCAA 145
```

## RESULT 3

```
US-10-023-208-58
; Sequence 58, Application US/10023208
; Publication No. US20030124537A1
; GENERAL INFORMATION:
; APPLICANT: Li, Min
; APPLICANT: Liu, Yuan-Ching
; TITLE OF INVENTION: PROCAROTIC LIBRARIES AND USES
; FILE REFERENCE: A-70174-1/RFT/RMS/RMK
; CURRENT APPLICATION NUMBER: US/10/023,208
; PRIOR FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US 60/256,163
; PRIOR FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 58
; LENGTH: 207
; TYPE: DNA
; ORGANISM: Artificial sequence
```

```
; FEATURE:
; OTHER INFORMATION: synthetic enzyme attachment site sequence
US-10-023-208-58
```

```
Query Match 75.1%; Score 107.4; DB 15; Length 207;
Best Local Similarity 85.1%; Pred. No. 2.4e-21;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
```

```
QY 3 ACCCTAGTATGAGTTGCCACTCCCTCTCTGCGGCTCGCTCGCTCGGTGGGCGCTG 62
Db 26 ACCCTAGTATGAGTTGCCACTCCCTCTCTGCGGCTCGCTCGCTCACTGAGCGCCG 85
QY 63 CGGACCAAGGTCGCGACGCGCAGAGCTCTGCTCTGCGGCGCCCGCCACCGCAGCGAGC 122
Db 86 GCGACCAAGGTCGCGCGACGCGCGGCTTTGCGCGCGGCGCTCACTGAGCGAGCGAGC 145
QY 123 GCGCAGAGAGGAGTGGGCAA 143
Db 146 GCGCAGAGAGGAGTGGCCAA 166
```

## RESULT 4

```
US-10-054-665-6
; Sequence 6, Application US/10054665
; Publication No. US20020197237A1
; GENERAL INFORMATION:
; APPLICANT: Engelhardt, John F.
; APPLICANT: Duan, Dongsheng
; TITLE OF INVENTION: Adeno-associated virus vectors
; FILE REFERENCE: 875.007US2
; CURRENT APPLICATION NUMBER: US/10/054,665
; PRIOR FILING DATE: 2002-06-13
; PRIOR APPLICATION NUMBER: US 09/276,625
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/086,166
; PRIOR FILING DATE: 1998-05-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 272
; TYPE: DNA
; ORGANISM: AAV circular intermediate, clone p1202
US-10-054-665-6
```

```
Query Match 75.1%; Score 107.4; DB 15; Length 272;
Best Local Similarity 85.1%; Pred. No. 2.3e-21;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
```

```
QY 3 ACCCTAGTATGAGTTGCCACTCCCTCTCTGCGGCTCGCTCGCTCGGTGGGCGCTG 62
Db 53 ACCCTAGTATGAGTTGCCACTCCCTCTCTGCGGCTCGCTCGCTCACTGAGCGCGG 112
QY 63 CGGACCAAGGTCGCGACGCGCAGAGCTCTGCTCTGCGGCGCCCGCCACCGCAGCGAGC 122
Db 113 GCGACCAAGGTCGCGCGACGCGCGGCTTTGTCGCGCGGCGCTCACTGAGCGAGCGAGC 172
QY 123 GCGCAGAGAGGAGTGGGCAA 143
Db 173 GCGCAGAGAGGAGTGGCCAA 193
```

## RESULT 5

```
US-09-782-378A-1
; Sequence 1, Application US/09782378A
; Patent No. US20020102731A1
; GENERAL INFORMATION:
; APPLICANT: Hearing, Patrick
; APPLICANT: Bahou, Wadie
; APPLICANT: Sandalon, Ziv
; APPLICANT: Gnatenko, Dmitri
; TITLE OF INVENTION: Adenoviral Vectors
; FILE REFERENCE: STONYB-04970
; CURRENT APPLICATION NUMBER: US/09/782,378A
```



```
; CURRENT FILING DATE: 2001-02-12
; PRIOR APPLICATION NUMBER: 60/237,747
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 4675
; TYPE: DNA
; ORGANISM: Human adeno-associated virus 2
US-09-782-378A-1
```

```
Query Match      75.1%; Score 107.4; DB 11; Length 4675;
Best Local Similarity 85.1%; Pred. No. 1.4e-21;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
```

```
OY      3 ACCCCTAGTATGATGAGTTGCCCACTCCCTCTCTGCGCGCTCGCTCGCTCGGTGGGCGCTG 62
      |||
DB      4535 ACCCCTAGTATGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGG 4594
OY      63 CGGACCAAGGTCGCCAGACGGCAGAGCTCTGCTTGCGCGCGCGCGCGCGCGCGCGCGCG 122
      |||
DB      4595 GCGACCAAGGTCGCCAGCGCCCGGGCTTTGCCCGGGCGGCTCACTGAGCGAGCGAGC 4654
OY      123 GCGCAGAGAGGAGTGGGCAA 143
      |||
DB      4655 GCGCAGAGAGGAGTGGCCAA 4675
```

## RESULT 6

```
US-09-782-378A-2
; Sequence 2, Application US/09782378A
; Patent No. US20020102731A1
; GENERAL INFORMATION:
; APPLICANT: Hearing, Patrick
; APPLICANT: Bahou, Wadie
; APPLICANT: Sandalon, Ziv
; APPLICANT: Gnatenko, Dmitri
; TITLE OF INVENTION: Adenoviral Vectors
; FILE REFERENCE: STONYB-04970
; CURRENT APPLICATION NUMBER: US/09/782,378A
; PRIOR FILING DATE: 2001-02-12
; PRIOR APPLICATION NUMBER: 60/237,747
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 4675
; TYPE: DNA
; ORGANISM: Human adeno-associated virus 2
US-09-782-378A-2
```

```
Query Match      75.1%; Score 107.4; DB 11; Length 4675;
Best Local Similarity 85.1%; Pred. No. 1.4e-21;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
```

```
OY      3 ACCCCTAGTATGATGAGTTGCCCACTCCCTCTCTGCGCGCTCGCTCGCTCGGTGGGCGCTG 62
      |||
DB      4535 ACCCCTAGTATGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGG 4594
OY      63 CGGACCAAGGTCGCCAGACGGCAGAGCTCTGCTTGCGCGCGCGCGCGCGCGCGCGCGCG 122
      |||
DB      4595 GCGACCAAGGTCGCCAGCGCCCGGGCTTTGCCCGGGCGGCTCACTGAGCGAGCGAGC 4654
OY      123 GCGCAGAGAGGAGTGGGCAA 143
      |||
DB      4655 GCGCAGAGAGGAGTGGCCAA 4675
```

```
RESULT 7
US-10-240-198-1
; Sequence 1, Application US/10240198
; Publication No. US20030100115A1
; GENERAL INFORMATION:
```

```
; APPLICANT: BTG International Ltd
; APPLICANT: BEARD DR, PETER
; APPLICANT: RAJ DR, KENNETH
; TITLE OF INVENTION: CYTOTOXIC AGENTS
; FILE REFERENCE: 142184WO
; CURRENT APPLICATION NUMBER: US/10/240,198
; CURRENT FILING DATE: 2002-09-30
; PRIOR APPLICATION NUMBER: 0009887.1
; PRIOR FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4675
; TYPE: DNA
; ORGANISM: adeno-associated virus 2
US-10-240-198-1
```

```
Query Match      75.1%; Score 107.4; DB 15; Length 4675;
Best Local Similarity 85.1%; Pred. No. 1.4e-21;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
```

```
OY      3 ACCCCTAGTATGATGAGTTGCCCACTCCCTCTCTGCGCGCTCGCTCGCTCGGTGGGCGCTG 62
      |||
DB      4535 ACCCCTAGTATGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGG 4594
OY      63 CGGACCAAGGTCGCCAGACGGCAGAGCTCTGCTTGCGCGCGCGCGCGCGCGCGCGCGCG 122
      |||
DB      4595 GCGACCAAGGTCGCCAGCGCCCGGGCTTTGCCCGGGCGGCTCACTGAGCGAGCGAGC 4654
OY      123 GCGCAGAGAGGAGTGGGCAA 143
      |||
DB      4655 GCGCAGAGAGGAGTGGCCAA 4675
```

## RESULT 8

```
US-09-804-898-1
; Sequence 1, Application US/09804898
; Patent No. US20020045264A1
; GENERAL INFORMATION:
; APPLICANT: DURING, MATTHEW
; APPLICANT: XIAO, WEIDONG
; TITLE OF INVENTION: PRODUCTION OF CHIMERIC CAPSID VECTORS
; FILE REFERENCE: 102182-14
; CURRENT APPLICATION NUMBER: US/09/804,898
; PRIOR FILING DATE: 2001-03-13
; PRIOR APPLICATION NUMBER: 60/189,110
; PRIOR FILING DATE: 2000-03-14
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 4679
; TYPE: DNA
; ORGANISM: adeno-associated virus 2
US-09-804-898-1
```

```
Query Match      75.1%; Score 107.4; DB 10; Length 4679;
Best Local Similarity 85.1%; Pred. No. 1.4e-21;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
```

```
OY      3 ACCCCTAGTATGATGAGTTGCCCACTCCCTCTCTGCGCGCTCGCTCGCTCGGTGGGCGCTG 62
      |||
DB      4539 ACCCCTAGTATGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGG 4598
OY      63 CGGACCAAGGTCGCCAGACGGCAGAGCTCTGCTTGCGCGCGCGCGCGCGCGCGCGCGCG 122
      |||
DB      4599 GCGACCAAGGTCGCCAGCGCCCGGGCTTTGCCCGGGCGGCTCACTGAGCGAGCGAGC 4658
OY      123 GCGCAGAGAGGAGTGGGCAA 143
      |||
DB      4659 GCGCAGAGAGGAGTGGCCAA 4679
```

## RESULT 9

```

US-09-945-681-10
; Sequence 10, Application US/09945681
; Patent No. US20020064878A1
; GENERAL INFORMATION:
; APPLICANT: UNIVERSITE DE NANTES
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR RECOMBINANT
; TITLE OF INVENTION: ADENO-ASSOCIATED VIRUS PRODUCTION
; FILE REFERENCE: B4182AA - UNIVERSITE DE NANTES
; CURRENT APPLICATION NUMBER: US/09/945,681
; CURRENT FILING DATE: 2001-09-05
; PRIOR APPLICATION NUMBER: PCT/EP 00/01854
; PRIOR FILING DATE: 2000-03-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 4679
; TYPE: DNA
; ORGANISM: adeno-associated virus 2
US-09-945-681-10

```

	Query Match	75.1%;	Score 107.4;	DB 10;	Length 4679;	
	Best Local Similarity	85.1%;	Pred. No. 1.4e-21;			
	Matches 120;	Conservative	0;	Mismatches 21;	Indels 0;	Gaps 0;
QY	3	ACCCCTAGTGTGAGTTGCCACTCCTCTCTGC	GCGCTCGTGCTCGGTGGGCTG	62		
Db	4539	ACCCTAGTATGATGGAGTTGGCCACTCCTCTCTG	CGCCTCGTGCTCACTGAGGCCG	4598		
QY	63	CGAACCAAAGTCCGCAGACGGCAGAGCTCTG	CTTGCCGGTCGCCACCGAGCGAGC	122		
Db	4599	GCCACCAAAGTTCGCCCGACGCCCGGCTTTG	CCCCGGCGGCTTAGTGAGCGAGCGAGC	4658		
QY	123	GCGCAGAAGAGGGAGTGGGCAA	143			
Db	4659	GCCCAGAGAGGGAGTGGCCAA	4679			

```

RESULT 10
US-10-038-972A-12
; Sequence 12, Application US/10038972A
; Publication No. US20020192823A1
; GENERAL INFORMATION:
; APPLICANT: J. Bartlett
; TITLE OF INVENTION: AAV VECTORS AND METHODS
; FILE REFERENCE: 28335/36996US
; CURRENT APPLICATION NUMBER: US/10/038,972A
; CURRENT FILING DATE: 2002-01-04
; PRIOR APPLICATION NUMBER: US 60/260,124
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 4679
; TYPE: DNA
; ORGANISM: adeno-associated virus 2
US-10-038-972A-12

```

Query Match	75.1%;	Score 107.4;	DB 15;	length 4679;
Best Local Similarity	85.1%;	Pred. No. 1.4e-21;		
Matches 120;	Conservative 0;	Mismatches 21;	Indels 0;	Gaps 0;
QY	3	ACCCCTAGTATGAGTGGCGCACTCCCTCTCTGCGCGCTCAGCTCGGTGGGCGCTG	62	
Db	4539	ACCCCTAGTATGAGTGGCGCACTCCCTCTCTGCGCGCTCAGCTCGGTGGGCGCG	4598	
QY	63	CGGACCAAGGTCCGAGACGGCAGAGCTCTGCTTGCCGGTCCACCGAGCGAGC	122	
Db	4599	GCGACCAAGGTCCGAGACGGCAGAGCTCTGCTTGCCGGTCCACCGAGCGAGC	4658	
QY	123	GCGCAGAGAGGAGTGGGCAA	143	
Db	4659	GCGCAGAGAGGAGTGGGCAA	4679	

RESULT 11  
US-10-077-294-1  
Sequence 1, Application US/10077294  
Publication No. US20020159979A1  
GENERAL INFORMATION:  
APPLICANT: Johnson, Philip R.  
TITLE OF INVENTION: Adeno-Associated Virus Materials and Methods  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 S. Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/077,294  
FILING DATE: 15-Feb-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/691,604  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: NO. US20020159979Aland, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 31975  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 474-6300  
TELEFAX: (312) 474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4680 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-10-077-294-1

```

Query Match      75.1%; Score 107.4; DB 15; Length 4680;
Best Local Similarity 85.1%; Pred. No. 1.4e-21;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

OY      3 ACCCCTAGTATGATGAGAGTTGCCCACTCCCTCTCTGTGCGCGCTGCTCGCTCGGTGGGGCCTG 62
      |||
      4540 ACCCCTAGTATGATGAGAGTTGCCCACTCCCTCTCTGTGCGCGCTGCTCGCTCACTGAGGCCGG 4599

OY      63 CGGACCAAGGTCGCCGACAGCGGCAGAGCTCTGTCTGCGGCGCCCAACGAGCGGAGCGAGC 122
      |||
      4600 GCGACCAAGGTCGCCCGACGCGCCGGGCTTTTGCCCGGCGGCGCTCACTGAGCGAGCGAGC 4659

DB      123 GCGCAGAGAGGGAGTGGGCCAA 143
      |||
      4660 GCGCAGAGAGGGAGTGGGCCAA 4680

RESULT 12
US-10-163-886-1
; Sequence 1, Application US/10163886
; Publication No. US20020187129A1
;
GENERAL INFORMATION:
;
APPLICANT: Johnson, Phillip R.
TITLE OF INVENTION: Adeno-Associated virus Materials and
Methods
;
NUMBER OF SEQUENCES: 3

```

;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
;; STREET: 6300 Sears Tower, 233 S. Wacker Drive  
;; CITY: Chicago  
;; STATE: Illinois  
;; COUNTRY: USA  
;; ZIP: 60606  
;;  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patentin Release #1.0, Version #1.25  
;;  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/10/163,886  
;; FILING DATE: 04-Jun-2002  
;; CLASSIFICATION: <Unknown>  
;;  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 09/292,703  
;; FILING DATE: <Unknown>  
;;  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: No. US20020187129A1and, Greta E.  
;; REGISTRATION NUMBER: 35,302  
;; REFERENCE/DOCKET NUMBER: 31975  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (312) 474-6300  
;; TELEFAX: (312) 474-0448  
;; TELEX: 25-3856  
;;  
;; INFORMATION FOR SEQ ID NO: 1:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 4680 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA (genomic)  
;; SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-10-163-886-1  
;  
Query Match 75.1%; Score 107.4; DB 15; Length 4680;  
Best Local Similarity 85.1%; Pred. No. 1.4e-21;  
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;  
;  
QY 3 ACCCTAGTATGAGTGGCCACTCCCTCTCTGCGGCTCGCTCGGTGGGCGCTG 62  
DB 4540 ACCCTAGTATGAGTGGCCACTCCCTCTCTGCGGCTCGCTCGGTGAGGCGCG 4599  
;  
QY 63 CGACCAAGGTCCGACAGCGGAGCTCTGCTGCGGCGGCCCAACGAGCGAGC 122  
DB 4600 GCGACCAAGGTCCGCGCGCGGCTTTGCCCGGCGCTCAGTGAGCGAGCGAGC 4659  
;  
QY 123 GCGCAGAGGAGGTGGCAA 143  
DB 4660 GCGCAGAGGAGGTGGCAA 4680  
;  
RESULT 13  
US-10-263-127-1  
;; Sequence 1, Application US/10263127  
;; Publication No. US20030082145A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Johnson, Philip R.  
;; TITLE OF INVENTION: Adeno-Associated Virus Materials and  
;; Methods  
;; NUMBER OF SEQUENCES: 3  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
;; STREET: 6300 Sears Tower, 233 S. Wacker Drive  
;; CITY: Chicago  
;; STATE: Illinois  
;; COUNTRY: USA  
;; ZIP: 60606  
;;  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: floppy disk  
;; COMPUTER: IBM PC compatible

;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patentin Release #1.0, Version #1.25  
;;  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/10/263,127  
;; FILING DATE: 02-Oct-2002  
;; CLASSIFICATION: <Unknown>  
;;  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US/10/077,294  
;; FILING DATE: 15-Feb-2002  
;; APPLICATION NUMBER: 09/691,604  
;; FILING DATE: <Unknown>  
;;  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: No. US20030082145A1and, Greta E.  
;; REGISTRATION NUMBER: 35,302  
;; REFERENCE/DOCKET NUMBER: 31975  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (312) 474-6300  
;; TELEFAX: (312) 474-0448  
;; TELEX: 25-3856  
;;  
;; INFORMATION FOR SEQ ID NO: 1:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 4680 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA (genomic)  
;; SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-10-263-127-1  
;  
Query Match 75.1%; Score 107.4; DB 15; Length 4680;  
Best Local Similarity 85.1%; Pred. No. 1.4e-21;  
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;  
;  
QY 3 ACCCTAGTATGAGTGGCCACTCCCTCTCTGCGGCTCGCTCGGTGGGCGCTG 62  
DB 4540 ACCCTAGTATGAGTGGCCACTCCCTCTCTGCGGCTCGCTCGGTGAGGCGCG 4599  
;  
QY 63 CGACCAAGGTCCGACAGCGGAGCTCTGCTGCGGCGGCCCAACGAGCGAGC 122  
DB 4600 GCGACCAAGGTCCGCGCGCGGCTTTGCCCGGCGCTCAGTGAGCGAGCGAGC 4659  
;  
QY 123 GCGCAGAGGAGGTGGCAA 143  
DB 4660 GCGCAGAGGAGGTGGCAA 4680  
;  
RESULT 14  
US-10-054-665-4  
;; Sequence 4, Application US/10054665  
;; Publication No. US20020197237A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Duan, Dongsheng  
;; TITLE OF INVENTION: Adeno-associated virus vectors  
;; FILE REFERENCE: 875.007US2  
;; CURRENT APPLICATION NUMBER: US/10/054,665  
;; CURRENT FILING DATE: 2002-06-13  
;; PRIOR APPLICATION NUMBER: US 09/276,625  
;; PRIOR FILING DATE: 1999-03-25  
;; PRIOR APPLICATION NUMBER: US 60/086,166  
;; PRIOR FILING DATE: 1998-05-20  
;; NUMBER OF SEQ ID NOS: 13  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 4  
;; LENGTH: 272  
;; TYPE: DNA  
;; ORGANISM: AAV circular intermediate, clone p81  
US-10-054-665-4  
;  
Query Match 74.0%; Score 105.8; DB 15; Length 272;  
Best Local Similarity 84.4%; Pred. No. 6.4e-21;  
Matches 119; Conservative 0; Mismatches 22; Indels 0; Gaps 0;



```

QY      3  ACCCTAGTATGAGTGTGCCACTCCCTCTCTGCGCGCTCGCTCGCTCGGTGGGCTG 62
      |||||||
Db      53  ACCCTAGTATGAGTGTGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGCGCG 112
      |||||||
QY      63  CGACCAAAAGTCCGACAGCGGACAGCTCTGCTGCGCGCCCAACCGAGCGAGCGAGC 122
      |||||||
Db      113  GCGCCAAAGTCCGCCAGCGCCCGGCTTGTCCCGGCGGCTCACTGAGCGAGCGAGC 172
      |||||||
QY      123  GCGCAGAGAGGAGTGGCAA 143
      |||||||
Db      173  GCGCAGAGAGGAGTGGCAA 193
      |||||||

```

RESULT 15  
US-10-054-665-6/c

```

; Sequence 6, Application US/10054665
; Publication No. US20020197237A1
; GENERAL INFORMATION:
; APPLICANT: Engelhardt, John F.
; APPLICANT: Duan, Dongsheng
; TITLE OF INVENTION: Adeno-associated virus vectors
; FILE REFERENCE: 875.007US2
; CURRENT APPLICATION NUMBER: US/10/054,665
; PRIOR APPLICATION NUMBER: US 09/276,625
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/086,166
; PRIOR FILING DATE: 1998-05-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 272
; TYPE: DNA
; ORGANISM: AAV circular intermediate, clone p1202
US-10-054-665-6

```

Query Match 71.7%; Score 102.6; DB 15; Length 272;  
Best Local Similarity 83.0%; Pred. No. 5e-20;  
Matches 117; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

```

QY      3  ACCCTAGTATGAGTGTGCCACTCCCTCTCTGCGCGCTCGCTCGCTCGGTGGGCTG 62
      |||||||
Db      209  ACCCTAGTATGAGTGTGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGCGCG 150
      |||||||
QY      63  CGACCAAAAGTCCGACAGCGGACAGCTCTGCTGCGCGCCCAACCGAGCGAGCGAGC 122
      |||||||
Db      149  GCGACCAAGCCCGGCGCTCGGCGACCTTGTGCGCCCGCTCACTGAGCGAGCGAGC 90
      |||||||
QY      123  GCGCAGAGAGGAGTGGCAA 143
      |||||||
Db      89  GCGCAGAGAGGAGTGGCAA 69
      |||||||

```

Search completed: July 20, 2003, 20:11:28  
Job time : 30.7204 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2003, 07:55:16 ; Search time 377.075 Seconds  
(without alignments)  
11036.790 Million cell updates/sec

Title: US-09-807-802A-1\_COPY\_4576\_4718  
Perfect score: 143  
Sequence: 1 ttaccctcagtgtgaggtc.....cgcagagaggagtggtgcaaa 143

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_mus:\*  
34: em\_htg\_pln:\*  
35: em\_htg\_rod:\*  
36: em\_htg\_mam:\*  
37: em\_htg\_vrt:\*  
38: em\_sy:\*  
39: em\_htgo\_hum:\*  
40: em\_htgo\_mus:\*  
41: em\_htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	143	100.0	4718	14	AF063497	AF063497 Adeno-ass
2	121.8	85.2	4683	14	AF028704	AF028704 Adeno-ass
3	119	83.2	4718	14	AF063497	AF063497 Adeno-ass
4	114.8	80.3	4683	14	AF028704	AF028704 Adeno-ass
5	113.2	79.2	4726	14	AV048704	U48704 Adeno-assoc
6	107.4	75.1	145	6	AR140333	AR140333 Sequence
7	107.4	75.1	145	6	AX286293	AX286293 Sequence
8	107.4	75.1	145	14	AA2LTR2	K01625 Adeno-assoc
9	107.4	75.1	165	6	AR034135	AR034135 Sequence
10	107.4	75.1	165	6	AX106702	AX106702 Sequence
11	107.4	75.1	165	6	I16806	I16806 Sequence 1
12	107.4	75.1	192	6	AA6400	AA6400 Sequence 3
13	107.4	75.1	198	6	AX418199	AX418199 Sequence
14	107.4	75.1	272	6	AX106701	AX106701 Sequence
15	107.4	75.1	4675	6	AX135805	AX135805 Sequence
16	107.4	75.1	4675	6	AX286292	AX286292 Sequence
17	107.4	75.1	4675	14	AA2CG	J01901 Adeno-assoc
18	107.4	75.1	4679	6	AX282480	AX282480 Sequence
19	107.4	75.1	4679	14	AF043303	AF043303 Adeno-ass
20	107.4	75.1	4680	6	AR028767	AR028767 Sequence
21	107.4	75.1	4680	6	I62303	I62303 Sequence 1
22	107.4	75.1	9258	12	AF383623	AF383623 Cloning v
23	107.4	74.0	191	6	AX398302	AX398302 Sequence
24	105.8	74.0	272	6	AX106699	AX106699 Sequence
25	105.8	74.0	272	6	AX106701	AX106701 Sequence
26	102.6	71.7	174	6	AX398304	AX398304 Sequence
27	101	70.6	145	6	AA6401	AA6401 Sequence 4
28	99.4	69.5	145	14	AA2LTR1	K01624 Adeno-assoc
29	99.4	69.5	145	14	AA2REPORT	M10681 Adeno-assoc
30	99.4	69.5	165	6	AR034135	AR034135 Sequence
31	99.4	69.5	165	6	AX106702	AX106702 Sequence
32	99.4	69.5	165	6	I16806	I16806 Sequence 1
33	99.4	69.5	2116	14	AA2LEFT	J01902 adeno-assoc
34	99.4	69.5	4675	6	AX135805	AX135805 Sequence
35	99.4	69.5	4675	6	AX286292	AX286292 Sequence
36	99.4	69.5	4675	14	AA2CG	J01901 Adeno-assoc
37	99.4	69.5	4679	6	AX282480	AX282480 Sequence
38	99.4	69.5	4680	6	AR028767	AR028767 Sequence
39	99.4	69.5	4680	6	AF043303	AF043303 Adeno-ass
40	99.4	69.5	4680	6	I62303	I62303 Sequence 1
41	99.4	69.5	8698	6	AX205072	AX205072 Sequence
42	99.4	69.5	9258	12	AF383623	AF383623 Cloning v
43	99.4	69.5	9258	14	AF028705	AF028705 Adeno-ass
44	98.8	69.1	4722	14	AV048704	U48704 Adeno-assoc
45	98.8	69.1	4726	14	AV048704	U48704 Adeno-assoc

ALIGNMENTS

RESULT 1  
AF063497 LOCUS 4718 bp DNA VRL 27-APR-1999  
DEFINITION Adeno-associated virus 1, complete genome.  
ACCESSION AF063497  
VERSION AF063497.1 GI:4689096  
KEYWORDS adeno-associated virus 1.  
SOURCE adeno-associated virus 1.  
ORGANISM Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.  
REFERENCE 1 (bases 1 to 4718)  
AUTHORS Xiao,W., Chirmule,N., Berta,S.C., McCullough,B., Gao,G. and Wilson,J.M.  
TITLE Gene therapy vectors based on adeno-associated virus type 1  
JOURNAL J. Virol. 73 (5), 3994-4003 (1999)

MEDLINE 99214338  
PUBMED 10196295  
REFERENCE 2 (bases 1 to 4718)  
AUTHORS Xiao,W. and Wilson,J.M.  
TITLE Direct Submission  
JOURNAL Submitted (05-MAY-1998) IHGT, University of Pennsylvania, 3601 Spruce Street, Philadelphia, PA 19104, USA  
FEATURES  
SOURCE  
1. .4718  
/organism="adeno-associated virus 1"  
/db\_xref="taxon:85106"  
/note="AAV1"  
335. .2206  
/codon\_start=1  
/product="nonstructural protein"  
/protein\_id="AAD27758.1"  
/db\_xref="GI:4689098"  
/translation="MPGFYEIVIKVPSDLDEHLPGISDSFVSWVAEKEMELPPDSMD LNLIEQAPLTVAEKLQRPDLVQMRVSKAPEALFFVQFEKESYFHLHILVETGVKS MVLGRFLSQIRDKLVQTIYRGIEPTLPNFAVTKTRNGAGGKRVDECYIPNYLLPK TOPELQWMTMMEBYISACINLAERKRLVAHDLTHVSQTOENKENTLPNSDAPVIRS KTSARYMELVGLVDRGITSEKQWIOEDQASYISFNAASNSRSQIKALDNAGKIMAL TKSAPDYLVGPAFPADIKTRIRYRIELNGYDPAYGVSFLGWAOKRFGKRTIWLFG PATGKTNIAEAIHAHVPFYGCNVNTNENFPNDQVDMVIMWEEGKMTAKVESAKA ILGSKRVVDOKCKSSAQIDPTPVTSTNTMCAVIDNSTTFEHOQPLQDRMFKEFL TRRLHDFGKVTQKEVKEFFRWAQDHTEVAHEFYVRKGGANKRPAPDDADKSEPKRA CPSVADPSTDAEGAPVDFADRYONKCSRHAGMLQMLPCKTCERNONFNICFTHGT RDCSECFPGVSESQPVVRKRTYRKLCAIHLLGRAPEIACSACDLVNVLDLDCVSEQ" 2223. .4433  
/codon\_start=1  
/product="capsid protein"  
/protein\_id="AAD27757.1"  
/db\_xref="GI:4689097"  
/translation="MAADGYLPDWLEDNLSGIREHWMWLKPGAPKPKANQOKODDGRG LVLPGYKYLGPFGNLGDKGEPVNAADAALHDKAYDQOLKAGDNPYLRYNHADAEPQE RLQEDTSFGNLAGRAVPAQAKRVLEPGLVEEGAKTAPGKKRPVEQSPOEPDSSGIG KTGQQPAKRLNFGQGTGDSSEVPDQPLGEPATPAVGPPTMASGGGAPMADNNEGA DGVNASGNMHCDSWLGDVITSTRTALPTYNNHLYKQISSASGASNDNHYFGY STPMGYFDENRFCHFSPPRDMQRLINNMWGFPRKLNFKLNIQVEVTTNDGVTTLA NMLSTVQVFSDEYQLPYVLGSAHQCLPPFPADVEMIPOGYLTLNNGSQAVGRSS FYCLEYFSPQMLRTGNNFTFSYTFEEVPHSSYAHQSOLDRLMNPILIDQLYLYNRTQ NQSGSAQNKDLFSRGSPPAGMSVQPKNMLPGCYRQQRVSKTKTDNNNSNFTWTGASK YNLNGRESIINPGTAMASHKDDKDKFPMSGVMIFGKESAGASNTALDNMTDEEEL KATNPVATERFGTVAVNFQSSSTDPATGVDHVMGALPGMWQDRDYILOGPIMAKIPH TDGHFHPSPLMGFGGLKHPPPQILIKNTVPANPAPFASATKPFITQYSTGQVSV EIMELQKENSKRWNPEVQYTSNYAKSANVDFTVDNGLYTEPRPIGTRYLTRPL" 1121 a 1393 c 1273 g 931 t  
BASE COUNT  
ORIGIN  
Query Match 100.0%; Score 143; DB 14; Length 4718;  
Best Local Similarity 100.0%; Pred. No. 3.4e-23;  
Matches 143; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TTACCCCTAGTGTGAGTGGCCCACTCCCTCTCTGCGCGCTCGCTCGCTCGGTGGGGCC 60  
Db 4576 TTACCCCTAGTGTGAGTGGCCCACTCCCTCTCTGCGCGCTCGCTCGGTGGGGCC 4635  
QY 61 TGGGACCAAGGTCGCGACAGCGGACAGAGCTCTGCTCTGCGCGCCGCCACCGAGCGGCGA 120  
Db 4636 TGGGACCAAGGTCGCGACAGCGGACAGAGCTCTGCTCTGCGCGCCGCCACCGAGCGGCGA 4695  
QY 121 GCGCGCAGAGAGGGAGTGGGCAA 143  
Db 4696 GCGCGCAGAGAGGGAGTGGGCAA 4718

SOURCE  
ORGANISM adeno-associated virus 6.  
REFERENCE 1 (bases 1 to 4683)  
AUTHORS Rutledge,E.A., Halbert,C.L. and Russell,D.W.  
TITLE Infectious clones and vectors derived from adeno-associated virus (AAV) serotypes other than AAV type 2  
JOURNAL J. Virol. 72 (1), 309-319 (1998)  
MEDLINE 98080418  
PUBMED 9420229  
REFERENCE 2 (bases 1 to 4683)  
AUTHORS Rutledge,E.A. and Russell,D.W.  
TITLE Direct Submission  
JOURNAL Submitted (24-SEP-1997) Hematology, Univ. of Washington, Box 357720, Seattle, WA 98195, USA  
FEATURES  
SOURCE  
1. .4683  
/organism="adeno-associated virus 6"  
/db\_xref="taxon:68558"  
/note="AAV6"  
320. .2191  
/codon\_start=1  
/evidence="not experimental"  
/product="nonstructural protein Rep78"  
/protein\_id="AAB95449.1"  
/db\_xref="GI:2766606"  
/translation="MPGFYEIVIKVPSDLDEHLPGISDSFVNWVAEKEMELPPDSMD LNLIEQAPLTVAEKLQRPDLVQMRVSKAPEALFFVQFEKESYFHLHILVETGVKS MVLGRFLSQIRDKLVQTIYRGIEPTLPNFAVTKTRNGAGGKRVDECYIPNYLLPK TOPELQWMTMMEBYISACINLAERKRLVAHDLTHVSQTOENKENTLPNSDAPVIRS KTSARYMELVGLVDRGITSEKQWIOEDQASYISFNAASNSRSQIKALDNAGKIMAL TKSAPDYLVGPAFPADIKTRIRYRIELNGYDPAYGVSFLGWAOKRFGKRTIWLFG PATGKTNIAEAIHAHVPFYGCNVNTNENFPNDQVDMVIMWEEGKMTAKVESAKA ILGSKRVVDOKCKSSAQIDPTPVTSTNTMCAVIDNSTTFEHOQPLQDRMFKEFL TRRLHDFGKVTQKEVKEFFRWAQDHTEVAHEFYVRKGGANKRPAPDDADKSEPKRA CPSVADPSTDAEGAPVDFADRYONKCSRHAGMLQMLPCKTCERNONFNICFTHGT RDCSECFPGVSESQPVVRKRTYRKLCAIHLLGRAPEIACSACDLVNVLDLDCVSEQ" 2208. .4418  
/codon\_start=1  
/evidence="not experimental"  
/product="capsid protein VP1"  
/protein\_id="AAB95450.1"  
/db\_xref="GI:2766607"  
/translation="MAADGYLPDWLEDNLSGIREHWMWLKPGAPKPKANQOKODDGRG LVLPGYKYLGPFGNLGDKGEPVNAADAALHDKAYDQOLKAGDNPYLRYNHADAEPQE RLQEDTSFGNLAGRAVPAQAKRVLEPGLVEEGAKTAPGKKRPVEQSPOEPDSSGIG KTGQQPAKRLNFGQGTGDSSEVPDQPLGEPATPAVGPPTMASGGGAPMADNNEGA DGVNASGNMHCDSWLGDVITSTRTALPTYNNHLYKQISSASGASNDNHYFGY STPMGYFDENRFCHFSPPRDMQRLINNMWGFPRKLNFKLNIQVEVTTNDGVTTLA NMLSTVQVFSDEYQLPYVLGSAHQCLPPFPADVEMIPOGYLTLNNGSQAVGRSS FYCLEYFSPQMLRTGNNFTFSYTFEDVPHSSYAHQSOLDRLMNPILIDQLYLYNRTQ NQSGSAQNKDLFSRGSPPAGMSVQPKNMLPGCYRQQRVSKTKTDNNNSNFTWTGASK YNLNGRESIINPGTAMASHKDDKDKFPMSGVMIFGKESAGASNTALDNMTDEEEL KATNPVATERFGTVAVNFQSSSTDPATGVDHVMGALPGMWQDRDYILOGPIMAKIPH TDGHFHPSPLMGFGGLKHPPPQILIKNTVPANPAPFASATKPFITQYSTGQVSV EIMELQKENSKRWNPEVQYTSNYAKSANVDFTVDNGLYTEPRPIGTRYLTRPL" 1114 a 1363 c 1277 g 929 t  
BASE COUNT  
ORIGIN  
Query Match 85.2%; Score 121.8; DB 14; Length 4683;  
Best Local Similarity 98.4%; Pred. No. 2.3e-18;  
Matches 123; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 19 TTGCCCACTCCCTCTCTGCGCGCTCGCTCGCTCGGTGGGGCTCGGACCAAGGTCCGC 78  
Db 4683 TTGCCCACTCCCTATAGCGCGCTCGCTCGCTCGGTGGGGCTCGGACCAAGGTCCGC 4624  
QY 79 AGACGGCAGAGCTCTGCTCTGCGCGCCGCCACCGAGCGAGCGCGCAGAGAGGAGTG 138  
Db 4623 AGACGGCAGAGCTCTGCTCTGCGCGCCGCCACCGAGCGAGCGCGCATAGAGGAGTG 4564  
QY 139 GGCAA 143



Db 4563 GCGAA 4559

RESULT 3

LOCUS AF063497/c 4718 bp DNA linear VRL 27-APR-1999

DEFINITION Adeno-associated virus 1, complete genome.

ACCESSION AF063497

VERSION AF063497.1 GI:4689096

KEYWORDS

SOURCE adeno-associated virus 1.

ORGANISM adeno-associated virus 1

REFERENCE 1 (bases 1 to 4718)

AUTHORS Xiao,W., Chirmule,N., Berta,S.C., McCullough,B., Gao,G. and Wilson,J.M.

TITLE Gene therapy vectors based on adeno-associated virus type 1

JOURNAL J. Virol. 73 (5), 3994-4003 (1999)

MEDLINE 99214338

PUBMED 10196295

REFERENCE 2 (bases 1 to 4718)

AUTHORS Xiao,W. and Wilson,J.M.

TITLE Direct Submision

JOURNAL Submitted (05-MAY-1998) IHGT, University of Pennsylvania, 3601 Spruce Street, Philadelphia, PA 19104, USA

FEATURES

source location/Qualifiers

1. 4718

/organism="adeno-associated virus 1"

/db\_xref="taxon:85106"

/note="AAV1"

335. .2206

/codon\_start=1

/product="nonstructural protein"

/protein\_id="AAD27758.1"

/db\_xref="GI:4689098"

/translation="MPGFYEIVIKVPSDLDEHLPGISDSFVNWVAEKEMELPPDSMD

LNLEQAPLTVAEKLQDRLVQWRVSKAPEALFVQFEKESYFHLHLVETTGVS

MVLGRFLSQIRDKLVQTIYRGIEPTLPNFWAVTKTRNGAGGKVVDECYIPNYLLPK

TOPELQAWMTNMEBYISACINAEKRLVAOHLTHVSOTQEQNKENLNPNSDAPYIRS

KTSARYMELVGLVDRGITSKQWIOEDQASISFNAASNSRSQIKALDNAGKTIAL

TKSAPDYLWGAPADIKTNRIYRILELNGYDPAVAGSVFLGWAQKRFGRNTIWLFG

PATGKTINIAEAIHAHVPFYGCNVNTNENFPENDCVDMKVIWEEGKMTAKVESAKA

ILGSKVRVDQCKCKSSAQIDPTPIVITSNTMCAVIDGNSITFEHQPLQDRMEKEEL

TRLEHDFGKVTQKEVKEFFRWQDHTVEVAHEFYVRKGANKRPAPDDADKSEPKRA

CPSVADPSTSDAEGAPVDPADRYQNKCSRHAGMLQMLFPCKTCERMNQNFNICFTHGT

RDCSECFPGVSESQPVVRKRTYRKLCAIHLLGRAPEIACSACDLVNVDLDDCVSEQ"

2223. .4433

/codon\_start=1

/product="capsid protein"

/protein\_id="AAD27757.1"

/db\_xref="GI:4689097"

/translation="MAADGYLPMWLEDNLSEGIREWMDLKGAPKPKANQOKODDGRG

LVLPGYKYLGFENGLDKGEFVNADAAALAEHDKAYDQOLKAGDNPYLRYNHADAEPQ

RLQEDTSFGNUGRAVFOAKKRVLEPGLVEEGAKTAPGKKRPVEQSPQEPDSSGIG

KTGOQPAKKRLNFGOTGDSSEVPDQPLGEPATPAVGPPTMASGGGAPMADNNEGA

DGVNAGSNWCHDSTWLGDRVITSTRTVALPTYNHLKYQISSASTGASNDNHYFGY

STPMGYFDNFRHCHFSPRDMORLINNNWGFPRKLNFKLNIQVKEVTINDGVTTIA

NNLSTVQVFSDEYQLPYVLGSAHQGLPPFPADVEMIPQGYLTLNNGSAVGRSS

FYCLEYFPSQMLRTGNNTFSYTEBEVPHSSYAHQSGLDRLMPLIDQYLYLNRTQ

NQSGSAQNKDLFSRGSFAGMSVQPKNMLPGPCYRQQVSKTKTDNNNSNFTWTGASK

YNLNGRESIINPGTAMASHKDDKDFPMSGVMIFGKESAGASNTALDNVMTDEEEL

KATNPVATERFGTAVANFOSSSTDPAIDVHAMGALPGMWQDRDVLQGPIMAKIPH

TDGHFHPSPLMGFGFLKNPPQILIKNTVPANPAEFASFTQYSTGQVSVE

IWEHJOKENSKRMNPQVQYTSNVAKSANDFTVDNNGLYTEBRPIGTRYLTRPL"

BASE COUNT 1121 a 1393 c 1273 g 931 t

ORIGIN

Query Match 83.2%; Score 119; DB 14; Length 4718;

Best Local Similarity 89.5%; Pred. No. 1e-17;

Matches 128; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 TTACCCCTAGTATGAGTGTGCCCACTCCCTCTGTGCGCGCTCGCTCGGTGGGCC 60

|||||

Db 143 TTACCCCTAGTATGAGTGTGCCCACTCCCTCTGTGCGCGCTCGCTCGGTGGGCC 84

QY 61 TCGGACCAAAAGGTCCGACAGCGCAGAGCTCTGCTCTGCGCGCCACCAGCAGCGCA 120

|||

Db 83 GGCAGACAGAGCTCTGCCGTGCGGACCTTTGTCGCGAGGCCACCAGCAGCGCA 24

QY 121 GCGCGACAGAGGAGTGGGCA 143

|||||

Db 23 GCGCGACAGAGGAGTGGGCA 1

RESULT 4

LOCUS AF028704 4683 bp DNA linear VRL 12-JAN-1998

DEFINITION Adeno-associated virus 6, complete genome.

ACCESSION AF028704

VERSION AF028704.1 GI:2766605

KEYWORDS

SOURCE adeno-associated virus 6.

ORGANISM adeno-associated virus 6

REFERENCE 1 (bases 1 to 4683)

AUTHORS Rutledge,E.A., Halbert,C.L. and Russell,D.W.

TITLE Infectious clones and vectors derived from adeno-associated virus (AAV) serotypes other than AAV type 2

JOURNAL J. Virol. 72 (1), 309-319 (1998)

MEDLINE 98080418

PUBMED 9420229

REFERENCE 2 (bases 1 to 4683)

AUTHORS Rutledge,E.A. and Russell,D.W.

TITLE Direct Submision

JOURNAL Submitted (24-SEP-1997) Hematology, Univ. of Washington, Box 357720, Seattle, WA 98195, USA

FEATURES

source location/Qualifiers

1. 4683

/organism="adeno-associated virus 6"

/db\_xref="taxon:68558"

/note="AAV6"

320. .2191

/codon\_start=1

/evidence=not\_experimental

/product="nonstructural protein Rep78"

/protein\_id="AAB95449.1"

/db\_xref="GI:2766606"

/translation="MEGFYEIVIKVPSDLDEHLPGISDSFVNWVAEKEMELPPDSMD

LNLEQAPLTVAEKLQDRLVQWRVSKAPEALFVQFEKESYFHLHLVETTGVS

MVLGRFLSQIRDKLVQTIYRGIEPTLPNFWAVTKTRNGAGGKVVDECYIPNYLLPK

TOPELQAWMTNMEBYISACINAEKRLVAOHLTHVSOTQEQNKENLNPNSDAPYIRS

KTSARYMELVGLVDRGITSKQWIOEDQASISFNAASNSRSQIKALDNAGKTIAL

TKSAPDYLWGAPADIKTNRIYRILELNGYDPAVAGSVFLGWAQKRFGRNTIWLFG

PATGKTINIAEAIHAHVPFYGCNVNTNENFPENDCVDMKVIWEEGKMTAKVESAKA

ILGSKVRVDQCKCKSSAQIDPTPIVITSNTMCAVIDGNSITFEHQPLQDRMEKEEL

TRLEHDFGKVTQKEVKEFFRWQDHTVEVAHEFYVRKGANKRPAPDDADKSEPKRA

CPSVADPSTSDAEGAPVDPADRYQNKCSRHAGMLQMLFPCKTCERMNQNFNICFTHGT

RDCSECFPGVSESQPVVRKRTYRKLCAIHLLGRAPEIACSACDLVNVDLDDCVSEQ"

2208. .4418

/codon\_start=1

/evidence=not\_experimental

/product="capsid protein VP1"

/protein\_id="AAB95450.1"

/db\_xref="GI:2766607"

/translation="MAADGYLPMWLEDNLSEGIREWMDLKGAPKPKANQOKODDGRG

LVLPGYKYLGFENGLDKGEFVNADAAALAEHDKAYDQOLKAGDNPYLRYNHADAEPQ

RLQEDTSFGNUGRAVFOAKKRVLEPGLVEEGAKTAPGKKRPVEQSPQEPDSSGIG

KTGOQPAKKRLNFGOTGDSSEVPDQPLGEPATPAVGPPTMASGGGAPMADNNEGA

DGVNAGSNWCHDSTWLGDRVITSTRTVALPTYNHLKYQISSASTGASNDNHYFGY

STPMGYFDNFRHCHFSPRDMORLINNNWGFPRKLNFKLNIQVKEVTINDGVTTIA

NNLSTVQVFSDEYQLPYVLGSAHQGLPPFPADVEMIPQGYLTLNNGSAVGRSS

FYCLEYFPSQMLRTGNNTFSYTEBEVPHSSYAHQSGLDRLMPLIDQYLYLNRTQ

NQSGSAQNKDLFSRGSFAGMSVQPKNMLPGPCYRQQVSKTKTDNNNSNFTWTGASK

YNLNGRESIINPGTAMASHKDDKDFPMSGVMIFGKESAGASNTALDNVMTDEEEL

KATNPVATERFGTAVANFOSSSTDPAIDVHAMGALPGMWQDRDVLQGPIMAKIPH

TDGHFHPSPLMGFGFLKNPPQILIKNTVPANPAEFASFTQYSTGQVSVE

```

BASE COUNT      1114 a      1363 c      1277 g      929 t
ORIGIN
Query Match      80.3%; Score 114.8; DB 14; Length 4683;
Best Local Similarity 88.0%; Pred. No. 9.3e-17;
Matches 125; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 2 TACCCCTAGTATGAGTTGCCCACTCCCTCTCTGCGCGCTCGCTCGGTGGGCGCT 61
    |||
DB 4542 TACCCCTAGTATGAGTTGCCCACTCCCTCTCTGCGCGCTCGCTCGGTGGGCGCG 4601

QY 62 GCGGACCAAGGTCCGACAGCGGACAGCTCTGCTCTGCGCGCGCGCGCGCGCGAG 121
    |||
DB 4602 GCAGAGCAGAGCTCTGCGCGCTCTGCGGACCTTTGCTCCGACGCGCGCGCGCGAG 4661

QY 122 GCGGACAGAGGAGTGCGGCAA 143
    |||
DB 4662 CGCGCATAGAGGAGTGCGCAA 4683

RESULT 5
AVU48704      4726 bp      DNA      linear      VRL 15-JUL-1996
LOCUS      Adeno-associated virus 3 nonstructural protein and capsid protein
DEFINITION      genes, complete cds, and complete genome.
ACCESSION      U48704
VERSION      U48704.1 GI:1408467
KEYWORDS
SOURCE      Adeno-associated virus 3 strain=3H.
ORGANISM      adeno-associated virus 3
REFERENCE      1 (bases 1 to 4726)
AUTHORS      Muramatsu, S., Mizukami, H., Young, N.S. and Brown, K.E.
TITLE      Nucleotide sequencing and generation of an infectious clone of
            adeno-associated virus 3
JOURNAL      Virology 221 (1), 208-217 (1996)
MEDLINE      96266430
PUBMED      8661429
REFERENCE      2 (bases 1 to 4726)
AUTHORS      Muramatsu, S. and Brown, K.E.
TITLE      Direct Submission
JOURNAL      Submitted (06-FEB-1996) Shin-ichi Muramatsu, Hematology Branch,
            NHLBI/NIH, 9000 Rockville, MD 20892, USA
FEATURES
    source
        1. .4726
            /organism="adeno-associated virus 3"
            /strain="3H"
            /db_xref="taxon:46350"
            /codon_start=1
            /product="nonstructural protein"
            /protein_id="AAC55048.1"
            /db_xref="GI:1408468"
            /translation="MPGFYEIVLKVPSLDLRLPGISNPFVNVAEKWDVPPDSMD
            PNLEQAPLTVAEKLQREFLVEMRVSKEBALFVQFEKGETVPHLVLIETIGVKS
            MNGRVYSQIKKLVIRIYRGVEPQLPNWFAVTKRNGAGGKAVDDCYIPNYLIPK
            TQPELQWMTNMDQYLASCTNLAEKRLVAQHLTHVSQTQEQNKENQNPNSDAVIRS
            KTSARYMELVGLVDRGITSSEKOWIOEASYSIPNAASNSRSQIKALDNASKIMSL
            TKTAPDYLVGSNPEDITKNRIYQILENGYDQYASVFLGNAQKFGKRTIMLFG
            PATTGKNTIAEAIAAHVPFYGCVMNTNENFPNDQVDMVIMWEGMTAKVESAKA
            ILGSKRVQCKSSAQIEPTPIVITSNTMCAVIDGNSITFEHQPLQDRMEFEL
            TRRLDHFQKVTQKQEKDFRWSADHTVAHEFYVRKGKGRASNDADVSEPKRE
            OTSLAQPTSDAEAPADYADRYQNKCSRHYGMNMLFPCKTCERNQOISNVCPTHGQR
            DCEGCEPQMSQPVSVKKTQYQKLPILHILGRAPEIACACLANVLDLDDCVSEQ
            "
        2209. .4419
            /codon_start=1
            /product="capsid protein"
            /protein_id="AAC55049.1"
            /db_xref="GI:1408469"
            /translation="MAADGYLPDWLEDNLSEGIRESMWALKPGVPOPKANQOQHODNRG
            LVLPGYKYLGPNGLDKGEPVNEADAAALEHDKAYDQQLKAGDNPYLKYNHADAERQE
            IEMELQENSKRWNPVEQYTSNYAKSANVDFVDNNGLYTEPRPIGTRYLTRPL"

```

```

RIQEDTSFGNLGRAVFOAKRILEPLGLVEEAKTAPGKKGAVDQSPQEPDSSGVG
KSGKQAPARKRLNFGQGTGDESVDPQPLGEPAPAPTSIGSNTMASGGAPMADNNEGA
DGVNSSGNMHCDSQWLDGRIYITSTRWALPTVNNHLYKQISSQSGASNDNHYEYS
TPWGYDFENRFCHFSFPRDMORLIINNMGFRPKLSFXLFNIQVRGVTQNDGTTIAN
NLSTVOVFTDSEYOLPYVLGSAHQGLPFPADVFMVPOGYLTLNNGSQAVERSSF
YCLEYFSPQMLRTGNMFPQSYTFREDVPFHSSYAHQSGLDRMLNPLIDQLYLNRTOG
TSGTNNOSRLIFSQAGPQMSLIQARNMLPGCYROORLSKTANDNNSNFPWTASK
YHNGRDSLVPNGPAMASHKDEEKFPPMHGNLIFEGEGTASNAELDNMTDEEEL
RTNPNVATEQYGTANNLOSSNTAPTGTVNHOALPGVMQDRDVLQGPIMAKIPH
TDGHEHPSPLMGFGFLKHPPEQIMIKNTPVPANPPTTSPAKFASFITQYSTGVSE
IEMELQENSKRWNPVEIQYTSNKNKSVNDFVDNNGVYSEPRPIGTRYLTRNL"

BASE COUNT      1200 a      1295 c      1251 g      980 t
ORIGIN
Query Match      79.2%; Score 113.2; DB 14; Length 4726;
Best Local Similarity 87.3%; Pred. No. 2.1e-16;
Matches 124; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 2 TACCCCTAGTATGAGTTGCCCACTCCCTCTCTGCGCGCTCGCTCGGTGGGCGCT 61
    |||
DB 4585 TACCTTAGTATGAGTTGGCCACTCCCTCTATGCGCACTCGCTCGGTGGGCGCT 4644

QY 62 GCGGACCAAGGTCCGACAGCGGACAGCTCTGCTCTGCGCGCGCGCGCGCGAG 121
    |||
DB 4645 GCGGACCAAGGTCCGACAGCGGACAGCTCTTGCACAGTCCGCGCGCGCGCGAG 4704

QY 122 GCGGACAGAGGAGTGCGCAA 143
    |||
DB 4705 TGCGCATAGAGGAGTGCGCAA 4726

RESULT 6
ARI40333/c
LOCUS      ARI40333      145 bp      DNA      linear      PAT 16-JUN-2001
DEFINITION      Sequence 1 from patent US 6207457.
ACCESSION      ARI40333
VERSION      ARI40333.1 GI:14482829
KEYWORDS
SOURCE      Unknown.
ORGANISM      Unknown.
REFERENCE      1 (bases 1 to 145)
AUTHORS      Natsoulis, G. and Surosky, R.T.
TITLE      Targeted nucleotide sequence delivery and integration system
JOURNAL      Patent: US 6207457-A 1 27-MAR-2001;
FEATURES
    source
        1. .145
            /organism="unknown"
            /product="nonstructural protein"
            /protein_id="AAC55048.1"
            /db_xref="taxon:46350"
            /codon_start=1
            /product="nonstructural protein"
            /protein_id="AAC55048.1"
            /db_xref="GI:1408468"
            /translation="MPGFYEIVLKVPSLDLRLPGISNPFVNVAEKWDVPPDSMD
            PNLEQAPLTVAEKLQREFLVEMRVSKEBALFVQFEKGETVPHLVLIETIGVKS
            MNGRVYSQIKKLVIRIYRGVEPQLPNWFAVTKRNGAGGKAVDDCYIPNYLIPK
            TQPELQWMTNMDQYLASCTNLAEKRLVAQHLTHVSQTQEQNKENQNPNSDAVIRS
            KTSARYMELVGLVDRGITSSEKOWIOEASYSIPNAASNSRSQIKALDNASKIMSL
            TKTAPDYLVGSNPEDITKNRIYQILENGYDQYASVFLGNAQKFGKRTIMLFG
            PATTGKNTIAEAIAAHVPFYGCVMNTNENFPNDQVDMVIMWEGMTAKVESAKA
            ILGSKRVQCKSSAQIEPTPIVITSNTMCAVIDGNSITFEHQPLQDRMEFEL
            TRRLDHFQKVTQKQEKDFRWSADHTVAHEFYVRKGKGRASNDADVSEPKRE
            OTSLAQPTSDAEAPADYADRYQNKCSRHYGMNMLFPCKTCERNQOISNVCPTHGQR
            DCEGCEPQMSQPVSVKKTQYQKLPILHILGRAPEIACACLANVLDLDDCVSEQ
            "
        2209. .4419
            /codon_start=1
            /product="capsid protein"
            /protein_id="AAC55049.1"
            /db_xref="GI:1408469"
            /translation="MAADGYLPDWLEDNLSEGIRESMWALKPGVPOPKANQOQHODNRG
            LVLPGYKYLGPNGLDKGEPVNEADAAALEHDKAYDQQLKAGDNPYLKYNHADAERQE
            IEMELQENSKRWNPVEQYTSNYAKSANVDFVDNNGLYTEPRPIGTRYLTRPL"

```

VERSION AX286293.1 GI:17048541  
KEYWORDS  
SOURCE adeno-associated virus 2.  
ORGANISM adeno-associated virus 2  
REFERENCE 1  
AUTHORS Raj,K. and Beard,P.M.  
TITLE Cytotoxic agents  
JOURNAL Patent: WO 0180840-A 2.01-NOV-2001;  
BTG INTERNATIONAL LIMITED (GB)  
FEATURES  
source 1. .145  
/organism="adeno-associated virus 2"  
/db\_xref="taxon:10804"  
misc\_structure 1. .145  
/note="ITR"  
misc\_feature 72  
/note="Unpaired base"  
misc\_feature 94  
/note="Unpaired base"  
BASE COUNT 24 a 49 c 52 g 20 t  
ORIGIN  
Query Match 75.1%; Score 107.4; DB 6; Length 145;  
Best Local Similarity 85.1%; Pred. No. 9.3e-15;  
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;  
QY 3 ACCCTAGTATGATGAGTTGCCCACTCCCTCTCTGCGCGCTCGCTCGCTGCGGCGCTG 62  
Db 5 ACCCTAGTATGATGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCGG 64  
QY 63 CGACCAAGGTCGCGACAGCGGACAGCTCTGCTTGGCGGCGGCGGCGGCGGCGGCGG 122  
Db 65 GCGACCAAGGTCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 124  
QY 123 GCGCAGAGAGGAGTGGGCAA 143  
Db 125 GCGCAGAGAGGAGTGGGCAA 145  
RESULT 8  
AA2LTR2 145 bp DNA linear VRL 27-APR-1993  
LOCUS Adeno-associated virus 2 right terminal sequence.  
DEFINITION K01625  
ACCESSION K01625.1 GI:209624  
VERSION K01625.1 GI:209624  
KEYWORDS replication; terminal repeat.  
SEGMENT 2 of 2  
SOURCE Adeno-associated virus 2H DNA (clone pSM620 [2]), from KB or HeLa cells.  
ORGANISM adeno-associated virus 2H  
REFERENCE 1 (bases 1 to 145)  
AUTHORS Lusby,E., Fife,K.H. and Berns,K.I.  
TITLE Nucleotide sequence of the inverted terminal repetition in adeno-associated virus DNA  
JOURNAL J. Virol. 34 (2), 402-409 (1980)  
MEDLINE 80185149  
PUBMED 6246271  
REFERENCE 2 (bases 1 to 145)  
AUTHORS Lefebvre,R.B., Riva,S. and Berns,K.I.  
TITLE Conformation takes precedence over sequence in adeno-associated virus DNA replication  
JOURNAL Mol. Cell. Biol. 4 (7), 1416-1419 (1984)  
MEDLINE 85061247  
PUBMED 6504049  
COMMENT The focus of both papers is the method of replication of the virus. [1] notes that the initial it is present only 30% of the time; it is shortened to t in 50% of the population and missing altogether in 15% of the population. There is further sequence heterogeneity which can be explained by assuming that the terminal 125 bases, which form an imperfect palindrome, are replaced by their inverted complement during replication. [2] found that deletion of the 9

terminal bases on the right and the 113 terminal bases on the left of AAV 2 genome did not stop DNA replication. Further deletion of an 11-base symmetrical sequence (bases 89 to 99) in the right terminal repetition inhibits DNA replication. Substitution of either an 8-base (cagatcg) or 12-base (cgcgatccg) symmetrical sequence unrelated to the original 11-base sequence restores DNA replication. All of this can be explained by assuming that the 125 base palindrome mentioned above form a t-shaped secondary structure which provides a primer for DNA polymerase during replication.  
FEATURES  
source 1. .145  
/organism="adeno-associated virus 2H"  
/db\_xref="taxon:10805"  
BASE COUNT 24 a 49 c 52 g 20 t  
ORIGIN 22 bases upstream of HaeIII sites.  
Query Match 75.1%; Score 107.4; DB 14; Length 145;  
Best Local Similarity 85.1%; Pred. No. 9.3e-15;  
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;  
QY 3 ACCCTAGTATGATGAGTTGCCCACTCCCTCTCTGCGCGCTCGCTCGCTGCGGCGCTG 62  
Db 5 ACCCTAGTATGATGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCGG 64  
QY 63 CGACCAAGGTCGCGACAGCGGACAGCTCTGCTTGGCGGCGGCGGCGGCGGCGGCGG 122  
Db 65 GCGACCAAGGTCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 124  
QY 123 GCGCAGAGAGGAGTGGGCAA 143  
Db 125 GCGCAGAGAGGAGTGGGCAA 145  
RESULT 9  
AR034135 165 bp DNA linear PAT 29-SEP-1999  
LOCUS Sequence 1 from patent US 5869305.  
DEFINITION AR034135  
ACCESSION AR034135  
VERSION AR034135.1 GI:5949740  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 165)  
AUTHORS Samulski,R.Jude. and Xiao,X.  
TITLE Recombinant viral vector system  
JOURNAL Patent: US 5869305-A 1 09-FEB-1999;  
FEATURES location/Qualifiers  
source 1. .165  
/organism="unknown"  
BASE COUNT 27 a 56 c 56 g 26 t  
ORIGIN  
Query Match 75.1%; Score 107.4; DB 6; Length 165;  
Best Local Similarity 85.1%; Pred. No. 9.1e-15;  
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;  
QY 3 ACCCTAGTATGAGTTGCCCACTCCCTCTCTGCGCGCTCGCTCGCTGCGGCGCTG 62  
Db 5 ACCCTAGTATGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCGG 64  
QY 63 CGACCAAGGTCGCGACAGCGGACAGCTCTGCTTGGCGGCGGCGGCGGCGGCGGCGG 122  
Db 65 GCGACCAAGGTCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 124  
QY 123 GCGCAGAGAGGAGTGGGCAA 143  
Db 125 GCGCAGAGAGGAGTGGGCAA 145  
RESULT 10  
AX106702 165 bp DNA linear PAT 30-APR-2001  
LOCUS AX106702



DEFINITION Sequence 7 from Patent WO0125465.  
ACCESSION AX106702  
VERSION AX106702.1 GI:13922363  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 165)  
AUTHORS Engelhardt, J.F., Dongsheng, D. and Ziyang, Y.  
TITLE Adeno-associated viruses and uses thereof  
JOURNAL Patent: WO 0125465-A 7 12-APR-2001;  
University of Iowa Research Foundation (US); Engelhardt, John F.  
(US); Dongsheng, Duan (US); Ziyang, Yan (US)  
location/Qualifiers  
1.165  
/organism="unidentified"  
/db\_xref="taxon:32644"  
/note="SEQ ID NO:1 of U.S. Patent No. 5,478,745"  
BASE COUNT 27 a 56 c 56 g 26 t  
ORIGIN  
Query Match 75.1%; Score 107.4; DB 6; Length 165;  
Best Local Similarity 85.1%; Pred. No. 9.1e-15;  
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;  
QY 3 ACCCTAGTATGAGTTGCCCACTCCCTCTCTGCGGCTGCTGCTCGTGGGGCCTG 62  
DB 5 ACCCTAGTATGAGTTGCCCACTCCCTCTCTGCGGCTGCTGCTCGTACTAGGCGG 64  
QY 63 CGACCAAGGTCGCGACGCGCAGAGCTCTGCTGCGGCGGCCCAACCGAGCGAGC 122  
DB 65 GCGACCAAGGTCGCGACGCGCGGCTTGGCGGCGGCTCAGTAGAGCGAGC 124  
QY 123 GCGCAGAGAGGAGTGGGCAA 143  
DB 125 GCGCAGAGAGGAGTGGCAA 145  
RESULT 11  
LOCUS 116806 165 bp DNA linear PAT 03-APR-1996  
DEFINITION Sequence 1 from patent US 5478745.  
ACCESSION 116806  
VERSION 116806.1 GI:1251714  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE Unclassified.  
1 (bases 1 to 165)  
AUTHORS Samulski, R.J. and Xiao, X.  
TITLE Recombinant viral vector system  
JOURNAL Patent: US 5478745-A 1 26-DEC-1995;  
location/Qualifiers  
1.165  
/organism="unknown"  
BASE COUNT 27 a 56 c 56 g 26 t  
ORIGIN  
Query Match 75.1%; Score 107.4; DB 6; Length 165;  
Best Local Similarity 85.1%; Pred. No. 9.1e-15;  
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;  
QY 3 ACCCTAGTATGAGTTGCCCACTCCCTCTCTGCGGCTGCTGCTCGTGGGGCCTG 62  
DB 5 ACCCTAGTATGAGTTGCCCACTCCCTCTCTGCGGCTGCTGCTCGTACTAGGCGG 64  
QY 63 CGACCAAGGTCGCGACGCGCAGAGCTCTGCTGCGGCGGCCCAACCGAGCGAGC 122  
DB 65 GCGACCAAGGTCGCGACGCGCGGCTTGGCGGCGGCTCAGTAGAGCGAGC 124  
QY 123 GCGCAGAGAGGAGTGGGCAA 143  
DB 125 GCGCAGAGAGGAGTGGCAA 145

RESULT 12  
LOCUS A46400 192 bp DNA linear PAT 07-MAR-1997  
DEFINITION Sequence 3 from Patent WO9523867.  
ACCESSION A46400  
VERSION A46400.1 GI:2300601  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 192)  
AUTHORS Denifle, P., Latta, M., Perricaudet, M. and Vigne, E.  
TITLE INTEGRATIVE RECOMBINANT ADENOVIRUSES, PREPARATION THEREOF AND THERAPEUTICAL USES THEREOF  
JOURNAL Patent: WO 9523867-A 3 08-SEP-1995;  
RHONE POULENC RORER SA (FR)  
Other publication AU 1852695 950918  
Other publication FR 2716893 950908.  
location/Qualifiers  
1.192.  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 39 a 58 c 63 g 32 t  
ORIGIN  
Query Match 75.1%; Score 107.4; DB 6; Length 192;  
Best Local Similarity 85.1%; Pred. No. 8.8e-15;  
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;  
QY 3 ACCCTAGTATGAGTTGCCCACTCCCTCTCTGCGGCTGCTGCTCGTGGGGCCTG 62  
DB 52 ACCCTAGTATGAGTTGCCCACTCCCTCTCTGCGGCTGCTGCTCGTACTAGGCGG 111  
QY 63 CGACCAAGGTCGCGACGCGCAGAGCTCTGCTGCGGCGGCCCAACCGAGCGAGC 122  
DB 112 GCGACCAAGGTCGCGACGCGCGGCTTGGCGGCGGCTCAGTAGAGCGAGC 171  
QY 123 GCGCAGAGAGGAGTGGGCAA 143  
DB 172 GCGCAGAGAGGAGTGGCAA 192  
RESULT 13  
LOCUS AX418199/c 198 bp DNA linear PAT 18-JUN-2002  
DEFINITION Sequence 1 from Patent WO0192551.  
ACCESSION AX418199  
VERSION AX418199.1 GI:21523203  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct.  
REFERENCE 1  
AUTHORS Samulski, R.J. and McCarty, D.M.  
TITLE Duplexed parvovirus vectors  
JOURNAL Patent: WO 0192551-A 1 06-DEC-2001;  
UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL (US)  
location/Qualifiers  
1.198  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="Inverted terminal repeat from the AAV-2 vector plasmid pSub 201"  
BASE COUNT 34 a 65 c 58 g 41 t  
ORIGIN  
Query Match 75.1%; Score 107.4; DB 6; Length 198;  
Best Local Similarity 85.1%; Pred. No. 8.7e-15;  
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;  
QY 3 ACCCTAGTATGAGTTGCCCACTCCCTCTCTGCGGCTGCTGCTCGTGGGGCCTG 62

Db 141 ACCCTAGATGAGTGGCCACTCCTCTCTGCGCGCTCGCTCACTAGAGCCGG 82  
QY 63 CGGACCAAGGTCCTGAGAGCGGAGAGTCTGCTGCGCGCGCGCGCGAGCGAGC 122  
Db 81 GCGACCAAGGTCGCGCGAGCGCGCGCTTGTGCGCGCGCTCACTAGAGCGAGC 22  
QY 123 GCGCAGAGGAGTGGCAA 143  
Db 21 GCGCAGAGGAGTGGCAA 1

RESULT 14

AX106701 272 bp DNA linear PAT 01-MAY-2001  
LOCUS AX106701  
DEFINITION Sequence 6 from Patent WO0125465.  
ACCESSION AX106701  
VERSION AX106701.1 GI:13922362  
KEYWORDS

SOURCE unidentified.  
ORGANISM unidentified.

REFERENCE 1 (bases 1 to 272)  
AUTHORS Engelhardt,J.F., Dongsheng,D. and Ziyang,Y.  
TITLE Adeno-associated viruses and uses thereof  
JOURNAL Patent: WO 0125465-A 6 12-APR-2001;  
University of Iowa Research Foundation (US) ; Engelhardt, John F.  
(US) ; Dongsheng, Duan (US) ; Ziyang, Yan (US)  
location/Qualifiers

FEATURES  
source 1..272  
/organism="unidentified"  
/db\_xref="taxon:32644"  
/note="AAV circular intermediate, clone p1202"

BASE COUNT 58 a 78 c 78 g 58 t  
ORIGIN

Query Match 75.1%; Score 107.4; DB 6; Length 272;  
Best Local Similarity 85.1%; Pred. No. 8.2e-15;  
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 3 ACCCTAGATGAGTGGCCACTCCTCTCTGCGCGCTCGCTCGGTGGGCGCTG 62  
Db 53 ACCCTAGATGAGTGGCCACTCCTCTCTGCGCGCTCGCTCACTAGAGCCGG 112  
QY 63 CGGACCAAGGTCCTGAGAGCGGAGAGTCTGCTGCGCGCGCGCGAGCGAGC 122  
Db 113 GCGACCAAGGTCGCGCGAGCGCGCGCTTGTGCGCGCGCTCACTAGAGCGAGC 172  
QY 123 GCGCAGAGGAGTGGCAA 143  
Db 173 GCGCAGAGGAGTGGCAA 193

RESULT 15

AX135805 4675 bp DNA linear PAT 29-MAY-2001  
LOCUS AX135805  
DEFINITION Sequence 5 from Patent WO0132711.  
ACCESSION AX135805  
VERSION AX135805.1 GI:14272040  
KEYWORDS

SOURCE adeno-associated virus 2.  
ORGANISM adeno-associated virus 2

REFERENCE 1 (bases,1 to 4675)  
AUTHORS Hermonat,P.L.  
TITLE Repo-associated virus aav rept8 major regulatory protein, mutants thereof and uses thereof  
JOURNAL Patent: WO 0132711-A 5 10-MAY-2001;  
THE BOARD OF TRUSTEES OF THE UNIVERSITY OF ARKANSAS (US)  
location/Qualifiers

FEATURES  
source 1..4675  
/organism="adeno-associated virus 2"  
/db\_xref="taxon:10804"

CDS

321..2186  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAC39573.1"  
/db\_xref="GI:14272041"  
/translation="MPGFYEIVIKVPSDLGHLPGISDSFVNMVAEKWELPPDSMD  
LNLIEQAPLTVAEKLRDFTLWEMRVSKAPEALFFVQFEKGSYFHHVAVETTVKS  
MVLGRFLSQIREKLIQRIYRGIEPTLPNWFVAVTKENAGGKGVNDECYIPNVLPR  
TQPELQWAMTWMEQYLISACNLTERKRLVAQHLTHVSQTQENKQNPNSDAPVIRS  
KTSARYMELVGLVDKGTSEKOWIOEDQASYISFNAAASNSRSQIKALDNGKMSL  
TKTAPDYLVGQPVEDISSNRIYKILELNGYDPQYAAVFLGWATKKFKRNTIWLFG  
PATGKTNIABAIATVPEYGVNMTNENFPNDGVDMVIMWEGKMTAKVESAKA  
ILGSKVRVDCKSSAQIDPTPVIIVTSNTNMCAYIDNSTTFEHOQPLQDRMFEEL  
TRRLDHDGKVTQKQEVKDFFRMAKDHVEVEHEFYKKGAKKRPAPSDADISEKRV  
RESVAQPSSTDAEASINYADRYQNKSRHVGMNMLFPCROCEMNONSNICTHGO  
DLECFPVSESQPVSVKAYOKLCYIHIMGVDPDACTADLVNVDLDDCTFEQ"

BASE COUNT 1198 a 1262 c 1251 g 964 t  
ORIGIN

Query Match 75.1%; Score 107.4; DB 6; Length 4675;  
Best Local Similarity 85.1%; Pred. No. 4.5e-15;  
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 3 ACCCTAGATGAGTGGCCACTCCTCTCTGCGCGCTCGCTCGGTGGGCGCTG 62  
Db 4535 ACCCTAGATGAGTGGCCACTCCTCTCTGCGCGCTCGCTCACTAGAGCCGG 4594  
QY 63 CGGACCAAGGTCCTGAGAGCGGAGAGTCTGCTGCGCGCGCGCGAGCGAGC 122  
Db 4595 GCGACCAAGGTCGCGCGAGCGCGCGCTTGTGCGCGCGCTCACTAGAGCGAGC 4654  
QY 123 GCGCAGAGGAGTGGCAA 143  
Db 4655 GCGCAGAGGAGTGGCAA 4675

Search completed: July 20, 2003, 13:19:54  
Job time : 378.075 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2003, 09:15:11 ; Search time 212.5 Seconds  
(without alignments)  
10898.630 Million cell updates/sec

Title: US-09-807-802A-1\_COPY\_4576\_4718

Perfect score: 143  
Sequence: 1 ttaccctagtgtgaggtc.....cgacagagggagtgggcaa 143

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 segs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*

1:	em_estba:*
2:	em_esthum:*
3:	em_estin:*
4:	em_estmu:*
5:	em_estov:*
6:	em_estpl:*
7:	em_estro:*
8:	em_hnc:*
9:	gb_est1:*
10:	gb_est2:*
11:	gb_hnc:*
12:	gb_est3:*
13:	gb_est4:*
14:	gb_est5:*
15:	em_estfun:*
16:	em_estom:*
17:	gb_gss:*
18:	em_gss_hum:*
19:	em_gss_inv:*
20:	em_gss_pln:*
21:	em_gss_vrt:*
22:	em_gss_fun:*
23:	em_gss_mam:*
24:	em_gss_mus:*
25:	em_gss_other:*
26:	em_gss_pro:*
27:	em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	37.8	26.4	763	BG490707	BG490707 602520211
c 2	37.4	26.2	644	BB632576	BB632576 BB632576
c 3	36.2	25.3	604	BI531460	BI531460 1024113H0
c 4	36.2	25.3	932	CNS0072Q	AL066742 Drosophila
c 5	35.6	24.9	578	BB651469	BB651469 BB651469
c 6	35.6	24.9	793	BI905802	BI905802 603062735

c 7	35.6	24.9	1115	14	BM906771	BM906771 AGENCOURT
c 8	35.4	24.8	1123	17	AG080476	AG080476 Pan trogl
c 9	35.2	24.6	8024	17	AG074763	AG074763 Pan trogl
c 10	35	24.5	844	17	CNS0052P	AL056652 Drosophila
c 11	34.6	24.2	860	12	BF578052	BF578052 602094705
c 12	34.4	24.1	890	17	AG127787	AG127787 Pan trogl
c 13	34.4	24.1	1136	13	BM416106	BM416106 OP21191 M
c 14	34.2	23.9	914	13	BI762701	BI762701 603048459
c 15	34.2	23.9	1009	17	CNS010EW	AL098882 Drosophila
c 16	34	23.8	623	10	BB613613	BB613613 BB613613
c 17	34	23.8	639	10	BB651977	BB651977 BB651977
c 18	34	23.8	643	10	BB652152	BB652152 BB652152
c 19	34	23.8	657	10	BB651054	BB651054 BB651054
c 20	34	23.8	661	10	BB621081	BB621081 BB621081
c 21	34	23.8	795	17	AQ331299	AQ331299 nbxb0049M
c 22	33.8	23.6	881	17	CNS031ZR	AL246240 Tetradon
c 23	33.8	23.6	932	17	CNS0072Q	AL066742 Drosophila
c 24	33.6	23.5	1372	13	BI913462	BI913462 603179075
c 25	33.6	23.5	1020	17	CNS004YJ	AL055391 Drosophila
c 26	33.6	23.5	1136	13	BM416106	BM416106 OP21191 M
c 27	33.4	23.4	468	9	AA043429	AA043429 zK54h10.s
c 28	33.4	23.4	655	13	BI907522	BI907522 603065665
c 29	33.4	23.4	683	13	BI768775	BI768775 603058268
c 30	33.4	23.4	891	13	BI906775	BI906775 603064520
c 31	33.4	23.4	1047	14	BM924938	BM924938 AGENCOURT
c 32	33.4	23.4	1284	13	BM547577	BM547577 AGENCOURT
c 33	33.2	23.2	283	9	AA874212	AA874212 vx03g11.r
c 34	33.2	23.2	879	17	AG141413	AG141413 Pan trogl
c 35	33	23.1	901	12	BF316571	BF316571 601903094
c 36	33	23.1	923	14	BQ718248	BQ718248 AGENCOURT
c 37	32.8	22.9	445	9	AL514067	AL514067 AL514067
c 38	32.8	22.9	925	17	CNS0091P	AL053013 Drosophila
c 39	32.6	22.8	298	17	CNS04M2X	AL296898 Tetradon
c 40	32.6	22.8	727	13	BI454382	BI454382 603170730
c 41	32.6	22.8	877	17	AG127610	AG127610 Pan trogl
c 42	32.6	22.8	984	12	BG809678	BG809678 mgct001xe
c 43	32.4	22.7	337	10	BE110673	BE110673 UI-R-BJ1-
c 44	32.4	22.7	787	17	AG133411	AG133411 Pan trogl
c 45	32.4	22.7	822	17	AG134283	AG134283 Pan trogl

#### ALIGNMENTS

RESULT 1  
LOCUS BG490707/c 763 bp mRNA linear EST 27-MAR-2001  
DEFINITION 602520211F1 NIH\_MGC\_18 Homo sapiens cDNA clone IMAGE:4638699 5', mRNA sequence.  
ACCESSION BG490707  
VERSION BG490707.1 GI:13452217  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 763)  
AUTHORS NIH-MGC http://mgs.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: DCTD/DTP/Gazdar  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: L1CM1401 row: e column: 04  
High quality sequence stop: 182.  
location/Qualifiers  
1. 763

FEATURES  
source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4638699"
/clone_1b="N1H MGC_18"
/tissue_type="large cell carcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
N1H MGC library."

```

BASE COUNT	188 a	241 c	310 g	24 c
ORIGIN				

Query Match	26.4%;	Score 37.8;	DB 12;	Length 763;
Best Local Similarity	58.4%;	Pred. No. 6.3;		
Matches 66;	Conservative 0;	Mismatches 47;	Indels 0;	Gaps 0;

QY 26 CTCCCTCTCTGCGCGCTCGCTCGCTGGGGGCTGCAGACCAAAAGTTCGAGACGGC 85  
| | | | | | | | | | | | | | | | | | |  
Db 261 CGCTGCGCTGCCCCGGGCGCTCGCTCGCTTGCCCTGCCCGCCCTCGCTTCGCGC 202

QY      86 AGAGCTTGTCTTGCCGCGCCACCGAGCGCAGCGGCCAGAGAGGGAGTG 138  
      ||| | | | | | | | | | | | | | |  
Db     201 AGCGCTACGCTTCGCGGCGCTGACCGCGCGAGGGA(C)GGCTGTGTCCGGCGTG 149

RESULT 2	BB632576	LOCUS	DEFINITION
BB632576	RIKEN full-length enriched, adult male hypothalamus Mus musculus cDNA clone A230067E15 5', mRNA sequence.	644 bp	mRNA linear EST 26-OCT-2001

ACCESSION	BB632576	
VERSION	BB632576.1	GI:16469164
KEYWORDS	EST.	
SOURCE	house mouse.	
ORGANISM	Mus musculus	

REFERENCE  
1 (bases 1 to 644)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

**AUTHORS**

Akakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toya, T., Muramatsu, M. and Hayashizaki, Y.

TITLE	RIKEN Mouse ESTs (Arakawa, T., et al. 2001)
JOURNAL	unpublished (2001)

COMMENT

Laboratory for Genome Exploration Research (Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

**Tel: 81-45-503-9222**  
**Fax: 81-45-503-9216**

Email: [genome-res@gsc.riken.go.jp](mailto:genome-res@gsc.riken.go.jp),  
URL: <http://genome.gsc.riken.go.jp/>

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length libraries for rapid discovery of "new

genes. Genome Res. : 10 (10), 1617-1630 (2000)

Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and

Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10(11):1555-1562 (2000)

Komno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y., and Hayashizaki, Y. Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. *Genome Res.* 11 (2), 281-289 (2001)

Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamanaka, I., Alizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and Hayashizaki, Y. Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences. *Mamm. Genome.* 12, 673-677 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

e mouse tissues.

FEATURES	Location/Qualifiers
source	1. .644

```

/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="A230067E15"
/clone_1b="RIKEN full-length enriched, adult male
hypothalmus"

```

```

/sex="male"
/tissue_type="hypothalamus"
/dev_stage="adult"
/lab_host="UDH10B"
/notes="Site 1: Sali; Site 2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGAGATCCACAGAGCTCTTTTTTTTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 20.0 and subtraction to Rot = 459.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGAGAGAGATTCTCAGTTAATTAAATTAAATCCCCCCCC
3']. cDNA was cleaved with XhoI and BamHI. Vector: a
modified pBluescript KS(+) after bulk excision from Lambda
FLC I. "
```

BASE	COUNT			
97	a	245	c	194
				g
				108
				t

Query Match	26.28;	Score 37.4;	DB 10;	Length 644;
Best Local Similarity	64.4%;	Pred. No. 7.5;		
Matches	56;	Conservative	0;	Mismatches 31;
			Indels	0;
			Gaps	0;

QY 29 CCTCTGCGCGCTCGCTCGCTGGTGGGGCCCTGCAGCAAAAGTTCGACAGCGGCAGA 88  
||||| ||||| | ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 535 CCTCTCGCGGCCCTGGCCAGGTGTTGGCCCCGGCGGGCAGAAAGACAAGCACGCGCATG 594

Qy	Db
89 GCTCTGCTCTGCGCGCCCGCCGAGCG	115
595 CCTCTACTTTGACGTGCCCCCGCGTG	621

RESULT 3  
BI531460/c

LOCUS	BI531460	604 bp	mRNA	linear	EST 29-AUG-
DEFINITION	1024113H09.y1 C. reinhardtii CC-1690, normalized, Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.				

ACCESSION	BI531460	1
VERSION	BI531460.1	GI:15372034

KEYWORDS	EST.
SOURCE	Chlamydomonas reinhardtii.

ORGANISM *Chlamydomonas reinhardtii*  
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;

REFERENCE  
1 (bases 1 to 604) Chlamydomonas; Chlamydomonas.

**AUTHORS** Grossman, A., Chang, C.-W., Davies, J., Harris, E., Hauser, C., Jelebvre, P., McDermott, J. P., Shrager, J., Silflow, C. and Stern, D.

TITLE Analyses of the *Chlamydomonas reinhardtii* Genome: A Model.













TITLE	Totoki, Y., Watanabe, H. and Sakaki, Y.		
JOURNAL	BAC end sequences of library PTB		
REFERENCE	Unpublished		
AUTHORS	2 (bases/1 to 890)		
TITLE	Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,		
JOURNAL	Totoki, Y., Watanabe, H. and Sakaki, Y.		
COMMENT	Direct Submission		
	Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical		
	and Chemical Research (RIKEN), Genomic Sciences Center (GSC),		
	1-7-22. Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan		
	(E-mail: chimbes@gsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,		
	Tel: 81-45-503-9111, Fax: 81-45-503-9170)		
	Clones are derived from the chimpanzee BAC library PTB This BAC end		
	was generated during the R&D process and may have higher chance of		
	clone tracking errors.		
	PRIMERS		
	Sequencing: -21M13		
LIBRARY	Vector : pKS145		
	R.Site 1 : SacI		
	R.Site 2 : SacI		
FEATURES	Location/Qualifiers		
source	1. .890		
	/organism="Pan troglodytes"		
	/db_xref="taxon:9598"		
	/clone="PTB-138M02.F"		
	/sex="male"		
	/cell_type="lymphoblast"		
	/clone_lib="PTB Chimpanzee Male BAC library"		
BASE COUNT	51 a 173 c 478 g 48 t 140 others		
ORIGIN			
Query Match	24.1%; Score 34.4; DB 17; Length 890;		
Best Local Similarity	57.3%; Pred. No. 50;		
Matches	59; Conservative 0; Mismatches 44; Indels 0; Gaps 0;		
OY	36 GCGCGCTCGCTCGCTCGGTGGGGCCCTGCGGACCAAGTCCGACAGCGGACGCTTGC 95		
Db	543 GCGCGGGGGCG 602		
OY	96 TCTGCCGCGCCCGCCAGCGAGCGAGCGCGCGCAGAGGAGTG 138		
Db	603 GCG 645		
RESULT 13			
LOCUS	BM416106 1136 bp mRNA linear EST 28-JAN-2002		
DEFINITION	OP21191 Mixed Stage EST's from Globodera pallida, the potato cyst		
ACCESSION	nematode Globodera pallida cDNA, mRNA sequence.		
VERSION	BM416106		
KEYWORDS	EST.		
SOURCE	Globodera pallida.		
ORGANISM	Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;		
	Tylenchoidea; Heteroderidae; Heteroderinae; Globodera.		
REFERENCE	1 (bases 1 to 1136)		
AUTHORS	Heer, J., Sosinski, B., Pokrzywa, R.M., Warry, A. and Opperman, C.		
TITLE	Mixed Stage EST's from Globodera pallida, the potato cyst nematode		
JOURNAL	Unpublished (2001)		
COMMENT	Contact: Opperman, C		
	Center for the Biology of Nematode Parasitism		
	NC State University; IACR-Rothamsted		
	Campus Box 7616; Raleigh, NC 27695, USA		
	Tel: 919.515.6699		
	Fax: 919.515.9500		
	Email: warthog@unity.ncsu.edu		
	GT11PCN1_D04_1-16R_030.ab1.		
FEATURES	Location/Qualifiers		
source	1. .1136		
	/organism="Globodera pallida"		
	/db_xref="taxon:36090"		

BASE COUNT	82 a	470 c	447 g	104 t	33 others
ORIGIN	/clone_lib="Mixed Stage EST's from Globodera pallida, the potato cyst nematode" /note="Vector: lambda GT11; This is a collaborative effort between IACR-Rothamsted and North Carolina State University. The library was constructed from mixed stage G. pallida in lambda GT11 by Paul Burroughs, IACR-Rothamsted."				
Query Match	24.1%; Score 34.4; DB 13; Length 1136;				
Best Local Similarity	58.1%; Pred. No. 54;				
Matches	79; Conservative 0; Mismatches 56; Indels 1; Gaps 1;				
QY	5 CCCTAGTATGAGATTGCCCACTCCCTCTTCGGCGCTCGCTCGGTGGGCTGCG 64				
Db	729 CCGCGGCGCGGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGG 670				
QY	65 GACCAAGGTCCGACAGCGGACAGAGCTCTGCTTGCAGCCGACCGAGCGAGCGC 124				
Db	669 GCGGGAAGGCCCG 611				
QY	125 GCAGAGAGGAGTGGG 140				
Db	610 GCGGGGGGGGGGGGGG 595				
RESULT 14	BI762701 914 bp mRNA linear EST 25-SEP-2001				
LOCUS	BI762701/c 603048459F1 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5188599 5',				
DEFINITION	mRNA sequence.				
ACCESSION	BI762701 GI:15754267				
VERSION	EST.				
KEYWORDS	human.				
SOURCE	Homo sapiens				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 914) NIH-MGC http://mgi.nci.nih.gov/.				
AUTHORS	National Institutes of Health, Mammalian Gene Collection (MGC)				
TITLE	Unpublished (1999)				
JOURNAL	Contact: Robert Strausberg, Ph.D.				
COMMENT	Email: cgabbs-r@mail.nih.gov				
	Tissue Procurement: Life Technologies, Inc.				
	cDNA Library Preparation: Life Technologies, Inc.				
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)				
	DNA Sequencing by: Incyte Genomics, Inc.				
	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:				
	http://image.llnl.gov				
	Plate: LLAM1471 row: e column: 16				
	High quality sequence stop: 859.				
FEATURES	Location/Qualifiers				
source	1. 914				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/clone="IMAGE:5188599"				
	/clone_lib="NIH_MGC_116"				
	/lab_host="DH10B"				
	/note="Organ: pooled colon, kidney, stomach; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA source anonymous pool of 3 colons, age 26 yo male, 49 yo female, 71 yo male colon; 46 yo male kidney, and pool of 2 stomachs, 62 yo male and 70 yo female. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.4 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 023. Note: this is a NIH MGC Library."				
BASE COUNT	155 a	292 c	275 g	192 t	



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2003, 07:55:16 ; Search time 377.075 Seconds  
(without alignments)  
11036.790 Million cell updates/sec

Title: US-09-807-802A-1\_COPY\_1\_143  
Perfect score: 143  
Sequence: 1 ttgccactccctctctg...aactccatcactaggggttaa 143

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 segs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:  
1: gb\_ba:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_cm:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vi:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_cm:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vi:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_mus:\*  
34: em\_htg\_pln:\*  
35: em\_htg\_rod:\*  
36: em\_htg\_mam:\*  
37: em\_htg\_vrt:\*  
38: em\_sy:\*  
39: em\_htgo\_hum:\*  
40: em\_htgo\_mus:\*  
41: em\_htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	143	100.0	4718	14	AF063497	AF063497 Adeno-ass
2	138.8	97.1	4683	14	AF028704	AF028704 Adeno-ass
3	119	83.2	4718	14	AF063497	AF063497 Adeno-ass
4	113.2	79.2	4722	14	AF028705	AF028705 Adeno-ass
5	113.2	79.2	4726	14	AVU48704	U48704 Adeno-assoc
6	112.2	78.5	4722	14	AF028705	AF028705 Adeno-ass
7	107.4	75.1	145	6	A46401	A46401 Sequence 4
8	107.4	75.1	145	14	AA2LTR1	K01624 Adeno-assoc
9	107.4	75.1	145	14	AA2REPORT	M10681 Adeno-assoc
10	107.4	75.1	165	6	AR034135	AR034135 Sequence
11	107.4	75.1	165	6	AX106702	AX106702 Sequence
12	107.4	75.1	165	6	I16806	I16806 Sequence 1
13	107.4	75.1	272	6	AX106701	AX106701 Sequence
14	107.4	75.1	2116	14	AA2LEFT	J01902 adeno-assoc
15	107.4	75.1	4675	6	AX135805	AX135805 Sequence
16	107.4	75.1	4675	6	AX286292	AX286292 Sequence
17	107.4	75.1	4675	14	AA2CG	J01901 Adeno-assoc
18	107.4	75.1	4679	6	AX282480	AX282480 Sequence
19	107.4	75.1	4679	14	AF043303	AF043303 Adeno-ass
20	107.4	75.1	4680	6	AR028767	AR028767 Sequence
21	107.4	75.1	4680	6	I62303	I62303 Sequence 1
22	107.4	75.1	4683	14	AF028704	AF028704 Adeno-ass
23	107.4	75.1	8698	6	AX205072	AX205072 Sequence
24	107.4	75.1	9258	12	AF383623	AF383623 Cloning v
25	105.8	74.0	272	6	AX106699	AX106699 Sequence
26	104.2	72.9	5585	6	AR054109	AR054109 Sequence
27	102.6	71.7	272	6	AX106701	AX106701 Sequence
28	99.4	69.5	145	6	AR140333	AR140333 Sequence
29	99.4	69.5	145	6	AX286293	AX286293 Sequence
30	99.4	69.5	145	14	AA2LTR2	K01625 Adeno-assoc
31	99.4	69.5	165	6	AR034135	AR034135 Sequence
32	99.4	69.5	165	6	AX106702	AX106702 Sequence
33	99.4	69.5	165	6	I16806	I16806 Sequence 1
34	99.4	69.5	192	6	A46400	A46400 Sequence 3
35	99.4	69.5	198	6	AX418199	AX418199 Sequence
36	99.4	69.5	4675	6	AX135805	AX135805 Sequence
37	99.4	69.5	4675	6	AX286292	AX286292 Sequence
38	99.4	69.5	4675	14	AA2CG	J01901 Adeno-assoc
39	99.4	69.5	4679	6	AX282480	AX282480 Sequence
40	99.4	69.5	4679	14	AF043303	AF043303 Adeno-ass
41	99.4	69.5	4680	6	AR028767	AR028767 Sequence
42	99.4	69.5	4680	6	I62303	I62303 Sequence 1
43	99.4	69.5	8698	6	AX205072	AX205072 Sequence
44	99.4	69.5	9258	12	AF383623	AF383623 Cloning v
45	98.8	69.1	4726	14	AVU48704	U48704 Adeno-assoc

## ALIGNMENTS

RESULT 1  
AF063497 LOCUS 4718 bp DNA linear VRL 27-APR-1999  
DEFINITION Adeno-associated virus 1, complete genome.  
ACCESSION AF063497  
VERSION AF063497.1 GI:4689096  
KEYWORDS adeno-associated virus 1.  
SOURCE adeno-associated virus 1.  
ORGANISM Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.  
REFERENCE 1 (bases 1 to 4718)  
AUTHORS Xiao, W., Chirmule, N., Berta, S.C., McCullough, B., Gao, G. and Wilson, J.M.  
TITLE Gene therapy vectors based on adeno-associated virus type 1  
JOURNAL J. Virol. 73 (5), 3994-4003 (1999)



MEDLINE 99214338  
 PUBMED 10196295  
 REFERENCE 2 (bases 1 to 4718)  
 AUTHORS Xiao, W. and Wilson, J.M.  
 TITLE Direct Submission  
 JOURNAL Submitted (05-MAY-1998) INGT, University of Pennsylvania, 3601  
 Spruce Street, Philadelphia, PA 19104, USA  
 FEATURES Location/Qualifiers

```
source 1. .4718
        /organism="adeno-associated virus 1"
        /db_xref="taxon:85106"
        /note="AAV1"
CDS 335. .2206
```

```

/codon_start=1
/product="nonstructural protein"
/protein_id="AAD2758.1"
/db_xref="GI:4689098"
/translation="MPGFYEIYIKVPSDLDEHLPGISDSFVSWVAEKEMELPPDSMD
LNLEIQAPLTVAEKLQDFDLVQMRRSKAPALFVQEGEKSEYFLHLIVETTVGKS
MYLGRFLSQIRDKLVQGTLYRGIEPTLLPNWFAVTKTNAGAGKVVDECIYNNLLPK
TQPELOWAWNTMEEYISACLNLAERKRLVAQHLTHVTSQTEQONKENLNPSDAPVIRRS
KTSARYMELVGMVLVDRGITSERQMIQEDQASYISFNAASNSRSQIKALDNAGKIMAL
PTSGADPYLVGPAPPAHADIAHVPFYGCNVWNTNENFPNDCYDKMVIWMEEGKMTAKVESKA
PATGSKTNIAEALIAHVPFYGCNVWNTNENFPNDCYDKMVIWMEEGKMTAKVESKA
ILGGSKVRVLDQCKSSAQIDIPVIVTSNTNMCALVDGNSTTFEHQQLPDMSEKFEI
TTRLEHDFGKVTQKEVKEFFRNAQDHYTEVAHEFYVRKGANRPAPDDAKSEPKRA
CPSPVADPSTSDAEGAPVDFADRIQNKCSRHAGMLNLPCKTCERNQNFNICFTHTG
RDCSECFPGVSESQPVVRKRTYRKLCAIHLLGRAPETACSAADLVNVDLDCVSEQ"
2223. .4433

```

```

/codon_start=1
/product="capsid protein"
/protein_id="AAD27757.1"
/db_xref="GI:4689097"
/translation="MAADGYLPDWLEDNLSSEGRHWMDLKPQAPKPKANQOKQDDGRG
LVLPQYKYLGPENGLDKEPVNAAADAAALKEADKAYDQOLKAGDNPYLYNHADAEPQE
RLQEDTSPFGNIGRAVFOAKKRVLPLGLEBAKTAAPGKTPRVEQSPQEPDSSGIG
KTGQQPAKKRLNFGQTGDSVDPDPLGEPATPAVGPRTMASGGGAPMADNNEGA
DGVGNASGNWHDSTLWLGDRVITTTSTRTWALPTYNHLLYKQISSASTGASNDNHFGY
STPMGYFDPNRPHCHFSPRDMQRLINNNNGRPRKLNFKLENIQVKEVTINDGVTTIA
NNLTSTVOVFSDEYQLPYVLGSAHOGCLPEPADVEMIPOYGYLTLNNGSOAVGRSS
FYCLEYFPSQMLRTGNNTFSTYFEEVPRHSSYAHOSQLDRMLNPLIDQYLYLYNRKT
NQSQAONKDLFSGRSAGMSVQPKNWLPGCPYQQRVSKTSKTDNNNSNFTWTGASK
YNLNGRESIINPGTAMASHKDEDEKFFPMGVMWIFKESAGASNTALDNVMTDEBEI
KATNPVATERFGTVAVNFQSSSTDPATGDVHAMGALLPGMWQDRDVLQGPIMAKIPH
TDGHFHPSPLMGGFGLKNPPOILLIKNTPVANPDAFSATKPAFSLTQYSTGQSVSE
IEWELQKENSKRWNPEVQYTSNYAKSANDVFTVDNNGLYTEPRPIGRTYLRPL"

```

	Query Match	100.0%;	Score 143;	DB 14;	length 4718;	
	Best Local Similarity	100.0%;	Pred. No. 5.7e-23;			
	Matches 143;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
OY	1 TTGCCACTCCCTCTCTGCGCGCTCGTCTGGTGGGGCTGCGGACCAAGTCGC	60				
Dp	1 TTGCCACTCCCTCTCTGCGCGCTCGTCTGGTGGGGCTGCGGACCAAGTCGC	60				
OY	61 AGACGGCAGAGCTCTGCTCTGCCGGCCCCACCAGCGAGCGCGCAGAGGGAGTG	120				
Dp	61 AGACGGCAGAGCTCTGCTCTGCCGGCCCCACCAGCGAGCGCGCAGAGGGAGTG	120				
OY	121 GGCAACTCCATCACTAAGGGTTAA	143				
Dp	121 GGCAACTCCATCACTAAGGGTTAA	143				

RESULT 2			
AF028704/c			
LOCUS	AF028704	4683 bp	DNA
DEFINITION	Adeno-associated virus 6, complete genome.		
ACCESSION	AF028704		
VERSION	AF028704.1		
KEYWORDS	GI:2766605		
			VRL 12-JAN-1998

SOURCE	adeno-associated virus 6.
ORGANISM	adeno-associated virus 6
REFERENCE	Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
AUTHORS	1 (bases 1 to 4683)
TITLE	Rutledge, E.A., Halbert, C.L. and Russell, D.W.
JOURNAL	Infectious clones and vectors derived from adeno-associated virus
MEDLINE	(AAV) serotypes other than AAV type 2
PUBMED	J. Virol. 72 (1), 309-319 (1998)
REFERENCE	98080418
AUTHORS	2 (bases 1 to 4683)
TITLE	Rutledge, E.A. and Russell, D.W.
JOURNAL	Direct Submission
FEATURES	Submitted (24-SEP-1997) Hematology, Univ. of Washington, Box
	357720, Seattle, WA 98195, USA
	Location/Qualifiers
	1. .4683

```

/organism="adeno-associated virus 6"
/db_xref="taxon:68558"
/note="AAV6"
320. .2191
CDS

```

```

/codon_start=1
/evidence=not_experimental
/product="nonstructural protein Rep78"
/protein_id="AAB95449.1"
/db_xref="GI:2766606"
/translation="MGFYEIVIKVPSDLDLHLPGISDSFVNVAEKEMELPPDSMD
LNLIEQAPLTVAEKIQDFLVQWRVRVSKAPEALFVQFEKESYFHLILVETGVKS
MVLGRFLSQIRDKLVOTIYRGIEPTLPNPAVTKTRNGAGGKAVDECYIPNYLLPK
TQPELQWAMTNMEEYISACLNLAERKLVADLLTVHSQTQEQNKENINPNSDAVIRS
KTSARMELVGMLVDRGITSEKQIQEDASYISFNASVRSQIKALDNAGKIMAL
TKSAPDYLVGPAPPADIKITNRIYRILENGYDPAVYAGSVFLGMAQKRGKNTIMLFG
PATGKTINIAEAIHAAVPEFYGCVMNTINENFPENDCVDKMIWMEEGKMTAKVESAKA
ILGSKVRVDQCKSSAQIDPTPIVITSNMCAVIDGNSTTFEHOPLQDRMFKFEL
TRLEHDFGKVTQKEVKEFERWAODHTEVAHEFYVRKGAGANKRPAADDKSEPKRA
CPSVADPSTSDAEGAPVDPADRYQNKCSRHAGMLQMLFPCKTCERNQNFNICFTHTGT
RDCSECFPGVSESQPPVVRKRTYRKLCAIHLLGRAPEIACSACDLVVVDLDDCVSEQ"
2208..4418

```

```

/codon_start=1
/evidence=not_experimental
/product="capsid protein VP1"
/protein_id="AAB95450.1"
/db_xref="GI:2766607"
/translation="MAADGYLPDWLEDNLSEGIREWMDLKPAPKPKANQOKQDDGRG
LVLPGYKYLGPENGLDKGEPAVNAADAALAEHDKAYDQOLKRGNNPYLRNHADAEFOE
RLQEDTSPFGGNLGRAVFOAKKRVLEPGLVEEGAKTAPGKKRPEVQSPEPDSSSGIG
KTGQQPAKKRLNFGQTGDSSEYVDPPQLGEPATPAAVGPTTMNPLIDQLYLYLNRTQ
DGVGNASGNWCHDSTWLGDRLVITSTRTVALPTYNHLXKQISSASTGASNDNHFYGY
STPMGWYEDENRFCHFSERPMQRLINNNWGRPKRLNFKLFNIQYKEVTTNDGVTITIA
NNLSTVQVESDEEYOLPYVLGSAHQGLPPPADVFMIPQGYLLTLNNGSOAVGRSS
FYCLEYFPQSÖMLRTGNNFTEYTFEDVPFHSSYAHSQSLDRLMNPLIDQLYLYLNRTQ
NQSGSAQNKDILLFSRSGPAGMSVQPKNWLPGPCYRQQRVSKTLDNNNSNFTWTGASK
YNLNGRESIINPGTAMASHKDXDKQKFFPMGSGVMI FGKESAGASNTALDNVMTDEEBEI
KATNPVATERFGTVAVNLQSSSTDPATGDVHMWGLPGMWQODRDVYLQGPIMAKIPH
TDGFHPSPILMGFGFLKHPPLIKNTPPANPAEFSAATKASFITQYSTGVQSVSE
IEMELQKENSKRWNPEVQYTSNVAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL"

```

Query Match	97.1%;	Score 138.8;	DB 14;	Length 4683;
Bed: Local Similarity	98.6%;	Pred. No. 5.1e-22;		
Matches 140;	Conservative 0;	Mismatches 2;	Indels 0;	Gaps 0;
OY	1	TTGCCCACTCCCTCTCTGCGCGCTCGCTCGCTCGGTGGGGCGCTGCGGACCAAGGTC	60	
Db	4683	TTGCCCACTCCCTCTATGCGCGCTCGCTCGGTGGGGCGCTGCGGACCAAGGTC	4624	
OY	61	AGACGGCAGAGCTCTGCTCTGCCCCGGCCCCACCGAGCGAGCGCGGCAGAGGGA	120	
Db	4623	AGACGGCAGAGCTCTGCTCTGCCCCGGCCCCACCGAGCGAGCGCGGCATGAGGGA	4564	
OY	121	GGCAACTCCATCACTAGGGGTA	142	

Db 4563 GGCAACTCCATCCTAGGGGTA 4542

RESULT 3

AF063497/c

LOCUS AF063497 4718 bp DNA linear VRL 27-APR-1999

DEFINITION Adeno-associated virus 1, complete genome.

ACCESSION AF063497

VERSION AF063497.1 GI:4689096

KEYWORDS

SOURCE adeno-associated virus 1.

ORGANISM adeno-associated virus 1

REFERENCE 1 (bases 1 to 4718)

AUTHORS Xiao, W., Chirmule, N., Berta, S.C., McCullough, B., Gao, G. and Wilson, J.M.

TITLE Gene therapy vectors based on adeno-associated virus type 1

JOURNAL J. Virol. 73 (5), 3994-4003 (1999)

MEDLINE 99214338

PUBMED 10196295

REFERENCE 2 (bases 1 to 4718)

AUTHORS Xiao, W. and Wilson, J.M.

TITLE Direct Submission

JOURNAL Submitted (05-MAY-1998) IHGT, University of Pennsylvania, 3601 Spruce Street, Philadelphia, PA 19104, USA

FEATURES

source location/Qualifiers

1. 4718

/organism="adeno-associated virus 1"

/db\_xref="taxon:85106"

/note="AAV1"

335. 2206

/codon\_start=1

/product="nonstructural protein"

/protein\_id="AAD2758.1"

/db\_xref="GI:4689098"

/translation="MPGFYEIVLKVPSPDLDEHLPGISDSFVSWVAEKEMWELPPDSMD LNLIEQAPLTVAEKLRQDFLVQRRVSKAPEALFVQFEKESYFHLHLVETTGYS MWLGRFLSQIRDKLVQTIYRGIEPTLPNMFVAVTKTRNGAGGNKVDECIYNYLLPK TQBELQWAMTNMEEYISACINLAERKLVQHLTHVSQTOENKKNLNPNSDAPVIRS KTSARYMELVGMVLDVDRGITSEKQWIOEDQASYISFNAASNSRSQIKALDNAGKIMAL TKSADPVLVGPAPADIKTRIRYRILENGYEPAYAGSVFLGWAOKRFGKRNITWIFG PATGKTNIAEAIHAHVPFYGCYVNTNENFPNDQVDMKVIWEEGKMTAKVESAKA ILGSKVRVDQCKSSAQIDPFIYVTSNTNMCVIDGNSTFEHQPLQDRMFEKEL TRRLDHFQKVTQKQVKEFFRMAQDHTVEVAHEFYVRKGAANKRPAPDDADKSEPKRA CPSVADPSTDAEGAPVDPADRYQNCISRHAGMLQMLPCKTCERMNQENICFTFGT RDSECFPGVSESQPVVRKRTYRKLCAIHLLGRAPEIACSACDLVNDLDDCVSEQ"

2223. 4433

/codon\_start=1

/product="capsid protein"

/protein\_id="AAD2757.1"

/db\_xref="GI:4689097"

/translation="MAADGYLPDWLEDNLSEGIREMMDLKPGVPOPKANQOHQDNRG LVLPGYKYLGPNGLDKGEPVNAADAALHDKAYDQOLKAGDNPYLRYNHADAEQOE RLQEDTSFGNLRGAVFOAKRVLLEPLGLVEGAKTAPGKKRPVEQSPOEPDSSSGVG KTGQOPAKKRLNFGQTDSESVDPQPLGEPAPATPAVGPPTMASGGGAPMADNNEGA DGVGNASGNMHCDSQWLGDRIYITSTRTVALPTYNNHLYKQISSASTGASNDNHYFGY STPMGYFDFNRFHCHFSPRDMQRLINNMGFRPKRLNFENIQVKEVTNDGVTTIA NNLSTVQVFTDSEYQLPVYLGSAHQGLPPPADVEMIPQYGLTLNNGSQAVGRSS FYCLEYFPSQMLRTGNNTFSYTFEEVPEHSSYASQSLDRMLNPLIDQLYLNRTO NQSGSAQNKDLFSRGSPPAGMSVQPKNWLPGPCYRQQRVSKTKTDNNSNFTWGTASK YNLNGRESIINPGTAMASHKDEBDEKFFPMGVMIFGKESAGASNTALDNVMTDEEIK ATNPVATERFGTAVANFQSSSTPATGVDHAMGALPGMWQDRDVLQGPIMAKIPH TDGHFHPSPLMGFGFLKHPPOQIMIKNTVPANPAEFSATKFASTQYSTGQVSE IEMELQENSKRMNPEVQYTSYNKSVANVDFTVDNNGLYTEPRPIGTRYLTRPL"

BASE COUNT 1121 a 1393 c 1273 g 931 t

ORIGIN

Query Match 83.2%; Score 119; DB 14; Length 4718;

Best Local Similarity 89.5%; Pred. No. 1.5e-17;

Matches 128; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 TTGCCACTCCCTCTCTGCGCGCTCGCTCGGTGGGGCGCTGCGGACCAAGTCCGC 60

|||||

Db 4718 TTGCCACTCCCTCTCTGCGCGCTCGCTCGCTCGGTGGGGCGGACAGACAGAGCTTTCG 4659

QY 61 AGACGGCAGAGCTCTGCTCTGCGCGCGCCACCGAGCGAGCGCGCAGAGGAGTGTG 120

Db 4658 CGTCTGCGGACCTTTGTCCTCGGAGGCCCCACCGAGCGAGCGCGCAGAGGAGTGTG 4599

QY 121 GGCACTCCATCCTAGGGGTAA 143

Db 4598 GGCACTCCATCCTAGGGGTAA 4576

RESULT 4

AF028705/c

LOCUS AF028705 4722 bp DNA linear VRL 12-JAN-1998

DEFINITION Adeno-associated virus 3B, complete genome.

ACCESSION AF028705

VERSION AF028705.1 GI:2766608

KEYWORDS

SOURCE adeno-associated virus 3B.

ORGANISM adeno-associated virus 3B

REFERENCE 1 (bases 1 to 4722)

AUTHORS Rutledge, E.A., Halbert, C.L. and Russell, D.W.

TITLE Infectious clones and vectors derived from adeno-associated virus (AAV) serotypes other than AAV type 2

JOURNAL J. Virol. 72 (1), 309-319 (1998)

MEDLINE 98080418

PUBMED 9420229

REFERENCE 2 (bases 1 to 4722)

AUTHORS Rutledge, E.A. and Russell, D.W.

TITLE Direct Submission

JOURNAL Submitted (25-SEP-1997) Hematology, Univ of Washington, Box 357720, Seattle, WA 98195, USA

FEATURES

source location/Qualifiers

1. 4722

/organism="adeno-associated virus 3B"

/db\_xref="taxon:68742"

/note="AAV3B"

317. 2191

/codon\_start=1

/product="nonstructural protein Rep78"

/protein\_id="AAB95451.1"

/db\_xref="GI:2766609"

/translation="MPGFYEIVLKVPSPDLDEHLPGISNSFVNVAEKEMWELPPDSMD PNLIEQAPLTVAEKLRQDFLVQRRVSKAPEALFVQFEKETYFHLHLVETTGYS MWLGRFLSQIRDKLVQTIYRGIEPTLPNMFVAVTKTRNGAGGNKVDDCIYNYLLPK TQBELQWAMTNMEOYISACINLAERKLVQHLTHVSQTOENKKNLNPNSDAPVIRS KTSARYMELVGMVLDVDRGITSEKQWIOEDQASYISFNAASNSRSQIKALDNAGKIMAL TKSADPVLVGSNPEDITKTRIRYRILENGYDPQYASVFLGWAOKRFGKRNITWIFG PATGKTNIAEAIHAHVPFYGCYVNTNENFPNDQVDMKVIWEEGKMTAKVESAKA ILGSKVRVDQCKSSAQIEPTPIYVTSNTNMCVIDGNSTFEHQPLQDRMFEKEL TRRLDHFQKVTQKQVKEFFRMAQDHTVEVAHEFYVRKGAANKRPASNDADVSEPKRO CTSLAQPSTSDAEPADYADRYQNCISRHAGMLQMLPCKTCERMNQISNVCFTFGOR DCEGCFPGMSESQPVVKKTKYQKLCPIHLLGRAPEIACSACDLANVLDLDDCVSEQ"

2208. 4418

/codon\_start=1

/product="capsid protein VP1"

/protein\_id="AAB95452.1"

/db\_xref="GI:2766610"

/translation="MAADGYLPDWLEDNLSEGIREMMDLKPGVPOPKANQOHQDNRG LVLPGYKYLGPNGLDKGEPVNEADAALHDKAYDQOLKAGDNPYLRYNHADAEQOE RLQEDTSFGNLRGAVFOAKRVLLEPLGLVEGAKTAPGKKRPVDSPOEPDSSSGVG KSGQOPAKKRLNFGQTDSESVDPQPLGEPAPATPAVGPPTMASGGGAPMADNNEGA DGVGNSSGNMHCDSQWLGDRIYITSTRTVALPTYNNHLYKQISSQSGASNDNHYFGYS TPMGYFDFNRFHCHFSPRDMQRLINNMGFRPKRLNFENIQVKEVTNDGVTTIAN NLTSTVQVFTDSEYQLPVYLGSAHQGLPPPADVEMIPQYGLTLNNGSQAVGRSSF XCLEYFPSQMLRTGNNTFSYTFEDVPEHSSYASQSLDRMLNPLIDQLYLNRTOG TTSGETNQSRLFSQAGPQSMISQARNWLPGPCYRQQRVSKTKTDNNSNFTWGTASK YHLNGRDSLVPNGPAMASHKDEBDEKFFPMHGNLIFGEGTTASNAELDNVMTDEEIK RTTNPVATEOYGTAVANLQSSNTAPTRITVNDGALPGMWQDRDVLQGPIMAKIPH TDGHFHPSPLMGFGFLKHPPOQIMIKNTVPANPAEFTTFSPAKFASTQYSTGQVSE IEMELQENSKRMNPEIQYTSYNKSVANVDFTVDNNGLYTEPRPIGTRYLTRNL"

BASE COUNT	1204 a	1297 c	1243 g	978 t
ORIGIN				
Query Match	79.2%; Score 113.2; DB 14; Length 4722;			
Best Local Similarity	87.3%; Pred. No. 3.2e-16;			
Matches 124; Conservative	0; Mismatches 18; Indels 0; Gaps 0;			
QY	1 TTGCCCACTCCCTCTCTGCGCGCTCGCTCGCTGGGGCCCTGCGACCAAGTCCGC 60			
DB	4722 TTGGCCACTCCCTCTATGCGCACTCGCTCGCTGGGGCCCTGCGACCAAGTCCGC 4663			
QY	61 AGACGGCAGAGCTCTGCTCTGCGCGCGCCCGCCAGCGAGCGAGCGGAGAGAGAGTGTG 120			
DB	4662 AGACGGAGCTGCTTGGACAGTCCGCGCCCGCCAGCGAGCGAGCGAGAGAGTGTG 4603			
QY	121 GGCACTCCATCACTAGGGGTA 142			
DB	4602 GCCAATCCATCACTAGAGGTA 4581			
RESULT 5				
LOCUS	AVU48704 4726 bp DNA linear VRL 15-JUL-1996			
DEFINITION	Adeno-associated virus 3 nonstructural protein and capsid protein genes, complete cds, and complete genome.			
ACCESSION	U48704			
VERSION	U48704.1 GI:1408467			
KEYWORDS				
SOURCE	Adeno-associated virus 3 strain=3H.			
ORGANISM	adeno-associated virus 3			
REFERENCE	1 (bases 1 to 4726)			
AUTHORS	Muramatsu, S., Mizukami, H., Young, N. S. and Brown, K. E.			
TITLE	Nucleotide sequencing and generation of an infectious clone of adeno-associated virus 3			
JOURNAL	Virology 221 (1), 208-217 (1996)			
MEDLINE	96266430			
PUBMED	8661429			
REFERENCE	2 (bases 1 to 4726)			
AUTHORS	Muramatsu, S. and Brown, K. E.			
TITLE	Direct Submission			
JOURNAL	Submitted (06-FEB-1996) Shin-ichi Muramatsu, Hematology Branch, NHLBI/NIH, 9000 Rockville, MD 20892, USA			
FEATURES	Location/Qualifiers			
source	1..4726			
	/organism="adeno-associated virus 3"			
	/strain="3H"			
	/db_xref="taxon:46350"			
	318..2192			
	/codon_start=1			
	/product="nonstructural protein"			
	/protein_id="AAC55048.1"			
	/db_xref="GI:1408468"			
	/translation="MPGFYEIVLKVPSDLDERLPGISNSFVNVAEKEWDVPPDSMD			
	PNLIEQAPLTVAEKLQREFLVEMRRVSKAPEALFVQFEKGETYFHLVLIETIGVKS			
	MVGGRVSGQIKEKLVTRIRYRGVEPOLPNMFVATKTRNGAGGKRVDDCYIPNYLPK			
	TQPELQWAWTNMDQYLSACINLAERKRLVAQHLTHVSQTOEQNKENQNPNSDAPVRS			
	KTSARVMELVGMLVDRGITSSEKQWIOEDQASYISFNASNSRSQIKALDNASKIMSL			
	TKTAPDYLVGSNPEDITKNRIYQILELNGYDPOYASVFLGWAQKKFGKRTIWLFG			
	PATGKTNIABAIHAHVPFGCVNWTNENFPNDQVDMVIMWEGKMTAKVESAKA			
	ILGSKVRVDQCKSSAQIEPTPIVTSNTMCAVIDGNTTFHQPLQDRMFKEFL			
	TRRLDHPFGKVTQKEVKDFFRWASDHVTDVAHEFYVRKGAKKRPASNDADVSEPKRQ			
	CTSLAQPTSDAEAPADYADRYONKCSRHVGNNMLFPCKTCERNNQISNVCFTHGOR			
	DCGECFPGMSESQPVSVKKTYQKLCPIHILGRAPEIACSDLANVLDLDCVSEQ			
	"			
	2209..4419			
	/codon_start=1			
	/product="capsid protein"			
	/protein_id="AAC55049.1"			
	/db_xref="GI:1408469"			
	/translation="MAADGYLPDWLEDNISGIRHWMALKPGVPQPKANQOHQDNRRG			
	LVLPGYKYLPGNGLDKGEFVNEADAALIEHDKAYDQOLKAGDNPYLKYNHADAIEFOE			
	RLQEDTISFGNIGRAVFOAKRILIEPLGLVEEAKTAPGKKGAVQSPQEPDSSSGVG			

BASE COUNT	1200 a	1295 c	1251 g	980 t
ORIGIN				
Query Match	79.2%; Score 113.2; DB 14; Length 4726;			
Best Local Similarity	87.3%; Pred. No. 3.2e-16;			
Matches 124; Conservative	0; Mismatches 18; Indels 0; Gaps 0;			
QY	1 TTGCCCACTCCCTCTCTGCGCGCTCGCTCGCTGGGGCCCTGCGACCAAGTCCGC 60			
DB	1 TTGGCCACTCCCTCTATGCGCACTCGCTCGCTGGGGCCCTGCGACCAAGTCCGC 60			
QY	61 AGACGGCAGAGCTCTGCTCTGCGCGCGCCCGCCAGCGAGCGAGCGGAGAGAGTGTG 120			
DB	61 AGACGGAGCTGCTTGGACAGTCCGCGCCCGCCAGCGAGCGAGCGAGAGAGTGTG 120			
QY	121 GGCACTCCATCACTAGGGGTA 142			
DB	121 GCCAATCCATCACTAGAGGTA 142			
RESULT 6				
LOCUS	AF028705 4722 bp DNA linear VRL 12-JAN-1998			
DEFINITION	Adeno-associated virus 3B, complete genome.			
ACCESSION	AF028705			
VERSION	AF028705.1 GI:2766608			
KEYWORDS				
SOURCE	adeno-associated virus 3B.			
ORGANISM	adeno-associated virus 3B			
REFERENCE	1 (bases 1 to 4722)			
AUTHORS	Rutledge, E. A., Halbert, C. L. and Russell, D. W.			
TITLE	Infectious clones and vectors derived from adeno-associated virus (AAV) serotypes other than AAV type 2			
JOURNAL	J. Virol. 72 (1), 309-319 (1998)			
MEDLINE	98080418			
PUBMED	9420229			
REFERENCE	2 (bases 1 to 4722)			
AUTHORS	Rutledge, E. A. and Russell, D. W.			
TITLE	Direct Submission			
JOURNAL	Submitted (25-SEP-1997) Hematology, Univ of Washington, Box 357720, Seattle, WA 98195, USA			
FEATURES	Location/Qualifiers			
source	1..4722			
	/organism="adeno-associated virus 3B"			
	/db_xref="taxon:68742"			
	/note="AAV3B"			
	317..2191			
	/codon_start=1			
	/product="nonstructural protein Rep78"			
	/protein_id="AAB95451.1"			
	/db_xref="GI:2766609"			
	/translation="MPGFYEIVLKVPSDLDEHLPGISNSFVNVAEKEWELPPDSMD			
	PNLIEQAPLTVAEKLQREFLVEMRRVSKAPEALFVQFEKGETYFHLVLIETIGVKS			
	MVGGRVSGQIKEKLVTRIRYRGVEPOLPNMFVATKTRNGAGGKRVDDCYIPNYLPK			
	TQPELQWAWTNMDQYLSACINLAERKRLVAQHLTHVSQTOEQNKENQNPNSDAPVRS			
	KTSARVMELVGMLVDRGITSSEKQWIOEDQASYISFNASNSRSQIKALDNASKIMSL			
	TKTAPDYLVGSNPEDITKNRIYQILELNGYDPOYASVFLGWAQKKFGKRTIWLFG			
	PATGKTNIABAIHAHVPFGCVNWTNENFPNDQVDMVIMWEGKMTAKVESAKA			
	ILGSKVRVDQCKSSAQIEPTPIVTSNTMCAVIDGNTTFHQPLQDRMFKEFL			
	TRRLDHPFGKVTQKEVKDFFRWASDHVTDVAHEFYVRKGAKKRPASNDADVSEPKRQ			
	CTSLAQPTSDAEAPADYADRYONKCSRHVGNNMLFPCKTCERNNQISNVCFTHGOR			
	DCGECFPGMSESQPVSVKKTYQKLCPIHILGRAPEIACSDLANVLDLDCVSEQ			



CDS  
2208..4418  
/codon\_start=1  
/product="capsid protein VP1"  
/protein\_id="AAB95452.1"  
/db\_xref="GI:2766610"  
/translation="MAADGYLPDWLEDNLSGIREWMALKPGVPQPKANQCHDNRRG  
LVLPGYKYLGPNGNGLDGEFVNEADAALEHDAADQOLKAGDNPLYLKNHDAEFOE  
RLQEDTSFGNGLGRAVFOAKRILEPLGLVEBAKTAPEGKRPVDSQPEPSSSGVG  
KSGKOPARKRLNFGCTGDSSEVDPDPLGEPAPAPTSIGNTMASGGGAPMADNNEGA  
DGVNSSGNWHCDQWLDGVITTTSTRWALPTYNHLYKQISSQSGASNDHYFGYS  
TPWGYFDENFRHCHFSPRDWRLINNMGFRPKLSFKLNIQVKEVTQNDGTTTIAN  
NLTSTVQVFTDSEYQLPYVLGSAHQGLPFPFADVFWVPQYGLTLNNGSQAVGRSSF  
YCLEYFSPQMLRTGNNPQSYTFEDVPFHSYAHQSILDRMLNPLIDQYLYLNRTQG  
TTSGETNQSRLFSQAGQMSLSQARNMLPGPCYQQLSKTANDNNSNFPMTAASK  
YHNGRDSLVPNGPAMASHKDEEKFPPMHGNIPEKEGTTASNAELDNMTDEEEI  
RTNPNVATBOYGTVANMLQSSNTAPTRTVNDQALPGMWQDRDVLQGPWAKIPH  
TDGHFSPPLMGFGELKHPPIQIMIKNTVPVAPNPPTTSPAKFASFITQYSTGVSE  
IEMELOKENSKRMNPEIQYTSNYSNVNDFTVDTNGVYSEPRPIGTRYLTRL"

BASE COUNT 1204 a 1297 c 1243 g 978 t

Query Match 78.5%; Score 112.2; DB 14; Length 4722;  
Best Local Similarity 87.2%; Pred. No. 5.4e-16;  
Matches 123; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 2 TGCCCACTCCCTCTCTGCGCGCTCGCTGCTCGTGGGGCTCGGACCAAGTCCGCA 61  
Db 1 TGGCACTCCCTCTATGCGCACTCGCTCGCTCGTGGGGCTCGGACCAAGTCCGCA 60

QY 62 GACGGCAGAGCTCTGCTTGCGCGCCCAACGAGCGAGCGAGCGCGCAGAGAGGAGTGG 121  
Db 61 GACGACGTGCTTGGCAGCTCCGCGCCCAACGAGCGAGCGAGCGAGTGCATAGAGGAGTGG 120

QY 122 GCACTCCATCACTAGGGTA 142  
Db 121 CCACTCCATCACTAGAGTA 141

RESULT 7  
A46401  
LOCUS A46401 145 bp DNA linear PAT 07-MAR-1997  
DEFINITION Sequence 4 from Patent WO9523867.  
ACCESSION A46401  
VERSION A46401.1 GI:2300602  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified  
REFERENCE 1 (bases 1 to 145)  
AUTHORS Deneffe, P., Latta, M., Perticaudet, M. and Vigne, E.  
TITLE INTEGRATIVE RECOMBINANT ADENOVIRUSES, PREPARATION THEREOF AND  
JOURNAL THERAPEUTICAL USES THEREOF  
COMMENT Patent: WO 9523867-A 4 08-SEP-1995;  
RHONE-POULENC RORER SA (FR)  
Other publication FR 2716893 950908.  
FEATURES  
source 1.145  
/organism="unidentified"  
/db\_xref="taxon:32644"

BASE COUNT 21 a 52 c 49 g 23 t

ORIGIN  
Query Match 75.1%; Score 107.4; DB 6; Length 145;  
Best Local Similarity 85.1%; Pred. No. 1.3e-14;  
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 1 TTGCCACTCCCTCTCTGCGCGCTCGCTGCTCGTGGGGCTCGGACCAAGTCCGC 60  
Db 1 TTGGCACTCCCTCTCTGCGCGCTCGCTGCTCACTGAGCGCGGCGACCAAGTCCGC 60

QY 61 AGACGGCAGAGCTCTGCTTGCGCGCCCAACGAGCGAGCGAGCGGCGCAGAGGAGTGTG 120

Db 61 CGACGCCCCGGCTTTGCCCCGGCCCTCACTGAGCGACGAGCGCGCAGAGAGGAGTGTG 120

QY 121 GGCACTCCATCACTAGGGGT 141  
Db 121 GGCACTCCATCACTAGGGGT 141

RESULT 8  
AA2LTR1  
LOCUS AA2LTR1 145 bp DNA linear VRL 27-APR-1993  
DEFINITION Adeno-associated virus 2 left terminal sequence.  
ACCESSION K01624  
VERSION K01624.1 GI:209623  
KEYWORDS replication; terminal repeat.  
SEGMENT 1 of 2  
SOURCE Adeno-associated virus 2H DNA, (clone pSM620 [2]), from KB or HeLa cells.  
ORGANISM adeno-associated virus 2H  
REFERENCE 1. (bases 1 to 145)  
AUTHORS Lusby, E., Fife, K.H. and Berns, K.I.  
TITLE Nucleotide sequence of the inverted terminal repetition in adeno-associated virus DNA  
JOURNAL J. Virol. 34 (2), 402-409 (1980)  
MEDLINE 80185149  
PUBMED 6246271  
REFERENCE 2 (bases 1 to 145)  
AUTHORS Lefebvre, R.B., Riva, S. and Berns, K.I.  
TITLE Conformation takes precedence over sequence in adeno-associated virus DNA replication  
JOURNAL Mol. Cell. Biol. 4 (7), 1416-1419 (1984)  
MEDLINE 85061247  
PUBMED 6504049  
COMMENT Both [1] and [2] present the opposite strand from the one presented here. The focus of both papers is the method of replication of the virus. [1] notes that the initial it is present only 30% of the time; it is shortened to t in 50% of the population and missing altogether in 15% of the population. There is further sequence heterogeneity which can be explained by assuming that the terminal 125 bases, which form an imperfect palindrome, are replaced by their inverted complement during replication. [2] found that deletion of the 9 terminal bases on the right and the 113 terminal bases on the left of AAV 2 genome did not stop DNA replication. Further deletion of an 11-base symmetrical sequence (bases 89 to 99) in the right terminal repetition inhibits DNA replication. Substitution of either an 8-base (cagatctg) or 12-base (cgggacccgg) symmetrical sequence unrelated to the original 11-base sequence restores DNA replication. All of this can be explained by assuming that the 125 base palindrome mentioned above form a t-shaped secondary structure which provides a primer for DNA polymerase during replication.  
FEATURES  
source 1.145  
/organism="adeno-associated virus 2H"  
/db\_xref="taxon:10805"

BASE COUNT 21 a 52 c 49 g 23 t

ORIGIN  
Query Match 75.1%; Score 107.4; DB 14; Length 145;  
Best Local Similarity 85.1%; Pred. No. 1.3e-14;  
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 1 TTGCCACTCCCTCTCTGCGCGCTCGCTGCTCGTGGGGCTCGGACCAAGTCCGC 60  
Db 1 TTGGCACTCCCTCTCTGCGCGCTCGCTGCTCACTGAGCGCGGCGACCAAGTCCGC 60

QY 61 AGACGGCAGAGCTCTGCTTGCGCGCCCAACGAGCGAGCGAGCGGCGCAGAGGAGTGTG 120  
Db 61 CGACGCCCCGGCTTTGCCCCGGCCCTCACTGAGCGAGCGGCGCAGAGGAGTGTG 120

QY 121 GGCACTCCATCACTAGGGGT 141

[illegible]

Db	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	FEATURES	BASE COUNT	ORIGIN	Query Match	Best Local Similarity	Matches
Db	81	CGACGCCCCGGCCTTTGCCCCGGGGCGCCCTCAGTACGACGAGCGGCGCAGAGAGGAGTG	140														
Qy	121	GGCACTCCATCACTAGGGGT	141														
Db	141	GCCCACTCCATCACTAGGGGT	161														
RESULT 11	AX106702	Sequence 7 from Patent WO0125465.	165 bp	DNA	linear	PAT 30-APR-2001											
LOCUS	AX106702																
DEFINITION	AX106702																
ACCESSION	AX106702																
VERSION	AX106702.1	GI:13922363															
KEYWORDS																	
SOURCE																	
ORGANISM																	
REFERENCE	1	(bases 1 to 165)															
AUTHORS	Engelhardt, J.F., Dongsheg, D. and Ziyang, Y.																
TITLE	Adeno-associated viruses and uses thereof																
JOURNAL	Patent: WO 0125465-A 7 12-APR-2001;																
	University of Iowa Research Foundation (US) ; Engelhardt, John F.																
	(US) ; Dongsheg, Duan (US) ; Ziyang, Yan (US)																
	Location/Qualifiers																
	1. .165																
	/organism="unidentified"																
	/db_xref="taxon:32644"																
	/note="SEO ID NO:1 of U.S. Patent No. 5,478,745"																
BASE COUNT	27 a	56 c	56 g	26 t													
ORIGIN																	
Query Match	75.1%;	Score 107.4;	DB 6;	Length 165;													
Best Local Similarity	85.1%;	Pred. No. 1.3e-14;															
Matches	120;	Conservative	0;	Mismatches	21;	Indels	0;	Gaps	0;								
Qy	1	TTGCCACTCCCTCTCTGCGCGCTCGCTCGCTGGGGCTGCGGACCAAGTCCGC	60														
Db	21	TTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCGCGGCGACCAAGTCCGC	80														
Qy	61	AGACGCGAGAGCTCTGCTCTGCGCGCGCGCCACGAGCGAGCGCGCAGAGGAGTG	120														
Db	81	CGACGCGCGCGCTTTGCCCGCGCGCGCTCAGTACGAGCGAGCGCGCAGAGGAGTG	140														
Qy	121	GGCACTCCATCACTAGGGGT	141														
Db	141	GCCCACTCCATCACTAGGGGT	161														
RESULT 12	AX106702	Sequence 1 from patent US 5478745.	165 bp	DNA</													

Db	21	TTGGCCACTCCCTCTGTGCGGCTCGTCTGCTCACTAGAGCGCGGCACCAAGTCGCC	80
OY	61	AGACGGCAGAGCTCTGCTCTGCCCCGCCACCAGCGAGCGCACCGCCAGAGAGGAGTG	120
Db	81	CGACGCCCCGGCTTTGCCCGGCGGCTCTAGTAGAGCGAGCGCGCAGAGAGGGAGTG	140
OY	121	GGAACAATCCATCACTAGGGGT	141
Db	141	GCCAATCATCACTAGGGGT	161
RESULT 13			
LOCUS	AX106701	272 bp	DNA linear PAT 01-MAY-2001
DEFINITION	Sequence 6 from Patent WO0125465.		
ACCESSION	AX106701		
VERSION	AX106701.1	GI:13922362	
KEYWORDS			
SOURCE	unidentified.		
ORGANISM	unclassified.		
REFERENCE	1 (bases 1 to 272)		
AUTHORS	Engelhardt,J.F., Dongsheng,D. and Ziyang,Y.		
TITLE	Adeno-associated viruses and uses thereof		
JOURNAL	Patent: WO 0125465-A 6 12-APR-2001; University of Iowa Research Foundation (US) ; Engelhardt, John F. (US) ; Dongsheng, Duan (US) ; Ziyang, Yan (US) location/Qualifiers 1..272 /organism="unidentified" /db_xref="taxon:32644" /note="AAV circular intermediate, clone p1202"		
BASE COUNT	58 a	78 c	78 g 58 t
ORIGIN			
Query Match 75.1%; Score 107.4; DB 6; Length 272; Best Local Similarity 85.1%; Pred.No.1.2e-14; Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;			
OY	1	TTGCCACTCCCTCTGTGCGGCTCGTCTGCTCGTGGGCGCTGCGACCAAGTCCGC	60
Db	69	TTGGCCACTCCCTCTGTGCGGCTCGTCTGCTCACTAGAGCGCGGCACCAAGTCCGC	128
OY	61	AGACGGCAGAGCTCTGCTCTGCCGCCGCCACCGAGCGAGCGCGCAGAGAGGAGTG	120
Db	129	CGACGCCCCGGCTTTGTCGCCCGGCTCTAGTAGAGCGAGCGCGCAGAGAGGGAGTG	188
OY	121	GGAACAATCCATCACTAGGGGT	141
Db	189	GCCAATCATCACTAGGGGT	209
RESULT 14			
LOCUS	AA2LEFT	2116 bp	DNA linear VRL 27-APR-1993
DEFINITION	adeno-associated virus 2 left half 45% of genome.		
ACCESSION	J01902		
VERSION	J01902.1	GI:209622	
KEYWORDS			
SOURCE	adeno-associated virus 2 from human hela cells.		
ORGANISM	adeno-associated virus 2		
REFERENCE	Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.		
AUTHORS	1 (bases 1 to 2116)		
TITLE	Lusby,E.W. and Berns,K.I. Mapping of the 5' termini of two adeno-associated virus 2 RNAs in the left half of the genome		
JOURNAL	J. Virol. 41 (2), 518-526 (1982)		
MEDLINE	82192580		
PUBMED	6281463		
FEATURES	Location/Qualifiers 1..2116 /organism="adeno-associated virus 2"		
Source			

	BASE COUNT	526 a	539 c	615 g	436 t	/db_xref="taxon:10804"
	ORIGIN					
	Query Match	75.1%;	Score 107.4;	DB 14;	Length 2116;	
	Best Local Similarity	85.1%;	Pred. No. 7.7e-15;			
	Matches 120; Conservative	0;	Mismatches 21;	Indels 0;	Gaps 0;	
OY	1 TTGCCACTCCCTCTCTGTGCGCGCTCGCTCGCTGGTGGGGCCTGCAGACCAAGGTCCG	60				
Db	1 TTGGCCACTCCCTCTCTGTGCGCGCTCGCTCGCTCACTGAGGCGCGGCGACCAAGGTCCG	60				
OY	61 AGACGGCAGAGCTCTGCTCTGCGCGCGCGCGCGCGCGCGCGCGAGCGAGGAGGAGTG	120				
Db	61 CGACGCCCCGGGCTTTGGCCCCGGCGCGCTCAGTAGAGCGAGCGCGCAGAGAGGAGTG	120				
OY	121 GGCACTCCATCACTAGGGGT 141					
Db	121 GCCAATCCATCACTAGGGGT 141					
	RESULT 15					
	AXI35805	4675 bp	DNA	linear	PAT 29-MAY-2001	
	LOCUS	Sequence 5 from Patent WO0132711.				
	DEFINITION	AXI35805				
	ACCESION	AXI35805				
	VERSION	AXI35805.1 GI:14272040				
	KEYWORDS	adeno-associated virus 2.				
	SOURCE	adeno-associated virus 2.				
	ORGANISM	Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.				
	REFERENCE	1 (bases 1 to 4675)				
	AUTHORS	Hermomat, P.L.				
	TITLE	Repo-associated virus aav rep78 major regulatory protein, mutants thereof and uses thereof				
	JOURNAL	Patent: WO 0132711-A 5 10-MAY-2001;				
	FEATURES	THE BOARD OF TRUSTEES OF THE UNIVERSITY OF ARKANSAS (US)				
	source	Location/Qualifiers				
	CDS	1..4675				
		/organism="adeno-associated virus 2"				
		/db_xref="taxon:10804"				
		321..2186				
		/note="unnamed protein product"				
		/codon_start=1				
		/protein_id="CAC39573.1"				
		/db_xref="GI:14272041"				
		/translation="MPGFYEIVIKVPSDL DGLPGISDFVNWAEEKWELPPDSDMD LNLIEQAPLVAEKLGRDFLETRRVSKAPEALFVQFEKGESYFNHVLVETGVKS MVLGRFLSQIREKLIQRIYRGIEPLPNWFAVTKTNGAGGKNVDECYIPNLLPK TPELOMAWTNMEQYLASACLNLTERKRLVAOHLTHVSQTQEONKENONPNSDAPIRS KTSARYMELVGMVLVDKGITSEKOWIOEQASYISFNASNSRSQIKALDNAGIMSL TKTAIDYLVGQQVEDISSNRITYKILELNGYDPQYAA SVFLGWATYKFGRNTIWLFG PATGKTNI AEALAH TVPFYGCVMNTNENFPENDCVDMVIWMEEGMTAKVESAKA ILGSKVRVDQCKSSAQIDPTPIVITSNTMCAVIDGNSTTFEHQOPLDRMFKEFL TRRLDHDFGKVTKEVKDFFRMAKHVEVEHEFYVKKGAKRPA PSDADISEPKRV RESVAQ PSTSDAEASINYADRYONCSRHVGMNLM LFCPCRCEMNONSNICFTHGOK DCLECPVSESQPVS VVKAYQLCYIHIMGKVDPACTACDLVNVLDLCIFEQ"				
	BASE COUNT	1198 a	1262 c	1251 g	964 t	
	ORIGIN					
	Query Match	75.1%;	Score 107.4;	DB 6;	Length 4675;	
	Best Local Similarity	85.1%;	Pred. No. 6.6e-15;			
	Matches 120; Conservative	0;	Mismatches 21;	Indels 0;	Gaps 0;	
OY	1 TTGCCACTCCCTCTCTGTGCGCGCTCGCTCGCTGGTGGGGCCTGCAGACCAAGGTCCG	60				
Db	1 TTGGCCACTCCCTCTCTGTGCGCGCTCGCTCGCTCACTGAGGCGCGGCGACCAAGGTCCG	60				
OY	61 AGACGGCAGAGCTCTGCTCTGCGCGCGCGCGCGCGCGCGCGAGCGAGGAGGAGTG	120				
Db	61 CGACGCCCCGGGCTTTGGCCCCGGCGCGCTCAGTAGAGCGAGCGCGCAGAGAGGAGTG	120				
OY	121 GGCACTCCATCACTAGGGGT 141					



Db 121 GCCAACTCCATCCTAGGGGT 141

Search completed: July 20, 2003, 13:19:53  
Job time : 377.075 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2003, 07:53:46 ; Search time 32.1207 Seconds  
(without alignments)  
10025.800 Million cell updates/sec

Title: US-09-807-802A-1\_COPY\_1\_143

Perfect score: 143  
Sequence: 1 ttgccactccctctctgcg.....aactccatcactaggsgttaa 143

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_101002:\*

1: /SIDS2/gcgdata/geneseq/geneqn-emb1/NA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneqn-emb1/NA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneqn-emb1/NA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneqn-emb1/NA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneqn-emb1/NA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneqn-emb1/NA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneqn-emb1/NA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneqn-emb1/NA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneqn-emb1/NA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneqn-emb1/NA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneqn-emb1/NA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneqn-emb1/NA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneqn-emb1/NA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneqn-emb1/NA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneqn-emb1/NA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneqn-emb1/NA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneqn-emb1/NA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneqn-emb1/NA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneqn-emb1/NA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneqn-emb1/NA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneqn-emb1/NA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneqn-emb1/NA2001A.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneqn-emb1/NA2001B.DAT:\*  
24: /SIDS2/gcgdata/geneseq/geneqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	143	100.0	143	21	AAD00781 Adeno-associated v
2	143	100.0	4718	21	AAD00772 Adeno-associated v
3	138.8	97.1	4683	22	AAF23749 AAV6 DNA sequence.
4	119	83.2	4718	21	AAD00772 Adeno-associated v
5	113.2	79.2	4722	22	AAF23748 AAV3B DNA sequence
6	112.2	78.5	4722	22	AAF23748 AAV3B DNA sequence
7	107.4	75.1	145	14	AAQ41448 AAV2 inverted term
8	107.4	75.1	145	16	AAT03385 Strict inverted te
9	107.4	75.1	145	20	AAK34295 Adeno-associated v

10	107.4	75.1	165	15	AAQ66769 Double-D ITR termi
11	107.4	75.1	165	18	AAT49462 Adeno associated v
12	107.4	75.1	165	22	AAD03535 Inverted terminal
13	107.4	75.1	272	21	AAZ47166 Head-to-tail ITR s
14	107.4	75.1	272	22	AAD03534 AAV circular inter
15	107.4	75.1	955	24	AAD37254 Adeno-associated v
16	107.4	75.1	955	24	AAD37254 Adeno-associated v
17	107.4	75.1	987	24	AAD37261 Adeno-associated v
18	107.4	75.1	987	24	AAD37261 Adeno-associated v
19	107.4	75.1	4414	24	AAD37260 Adeno-associated v
20	107.4	75.1	4414	24	AAD37260 Adeno-associated v
21	107.4	75.1	4476	24	AAD37259 Adeno-associated v
22	107.4	75.1	4476	24	AAD37259 Adeno-associated v
23	107.4	75.1	4498	24	AAD37258 Adeno-associated v
24	107.4	75.1	4498	24	AAD37258 Adeno-associated v
25	107.4	75.1	4675	22	AAH41481 Adeno-associated v
26	107.4	75.1	4675	22	AAH89931 Nucleotide sequenc
27	107.4	75.1	4675	24	ABA02989 Adeno-associated v
28	107.4	75.1	4679	22	AA166974 Adeno-associated v
29	107.4	75.1	4679	22	AAF23750 AAV2 DNA sequence.
30	107.4	75.1	4680	17	AAT09008 Wild-type adeno-as
31	107.4	75.1	4683	22	AAF23749 AAV6 DNA sequence.
32	107.4	75.1	4825	24	AAD37257 Adeno-associated v
33	107.4	75.1	4825	24	AAD37257 Adeno-associated v
34	107.4	75.1	4848	24	AAD37263 Adeno-associated v
35	107.4	75.1	4966	24	AAD37256 Adeno-associated v
36	107.4	75.1	4966	24	AAD37256 Adeno-associated v
37	107.4	75.1	4990	24	AAD37262 Adeno-associated v
38	107.4	75.1	4990	24	AAD37262 Adeno-associated v
39	107.4	75.1	5060	24	AAD37264 Adeno-associated v
40	107.4	75.1	5060	24	AAD37264 Adeno-associated v
41	107.4	75.1	5149	24	AAD37255 Adeno-associated v
42	107.4	75.1	5149	24	AAD37255 Adeno-associated v
43	107.4	75.1	5932	21	AAZ45928 Nucleotide sequenc
44	107.4	75.1	5932	21	AAZ45928 Nucleotide sequenc
45	107.4	75.1	6142	21	AAZ45932 Nucleotide sequenc

ALIGNMENTS

RESULT 1	
AAD00781	
ID	AAD00781 standard; DNA; 143 BP.
XX	
AC	AAD00781;
XX	
XX	
DT	08-SEP-2000 (first entry)
XX	
DE	Adeno-associated virus serotype 1 5' inverted terminal repeat.
XX	
KW	Adeno-associated virus serotype 1; AAV-1; rep protein; capsid protein;
KW	cap protein; recombinant viral vector; gene delivery; gene therapy;
KW	vaccine; transgene; inverted terminal repeat; ITR; ss.
XX	
OS	Adeno associated virus serotype 1.
XX	
FH	
FT	Key
FT	misc_binding
FT	Location/Qualifiers
FT	1..41
FT	/tag= a
FT	/bound_moiety= "binds to nucleotides 85..125"
FT	42..62
FT	/tag= b
FT	64..84
FT	/tag= c
FT	85..125
FT	/tag= d
FT	/bound_moiety= "binds to nucleotides 41..1"
XX	
PN	WO200028061-A2.
XX	
PD	18-MAY-2000.
XX	

PF 02-NOV-1999; 99WO-US25694.  
XX  
PR 05-NOV-1998; 98US-0107114.  
XX  
PA (UYPE-) UNIV PENNSYLVANIA.  
XX  
PI Wilson JM, Xiao W;  
XX  
DR WPI; 2000-376571/32.  
XX  
PT Novel adeno-associated virus serotype 1 polynucleotide useful for  
PT preparation of medicament for delivery of a transgene to a host -  
XX  
PS Example 2; Fig 2; 108pp; English.  
XX  
CC The patent discloses an adeno-associated virus serotype 1 (AAV-1) DNA  
CC which is characterised by two inverted terminal repeats (ITR) and open  
CC reading frames for rep and capsid (cap) proteins. The rep reading frame  
CC encodes four proteins, Rep 78, Rep 68, Rep 52 and Rep 40, while the cap  
CC reading frame encodes three structural proteins, VP1, VP2 and VP3.  
CC The AAV-1 sequence or its fragments particularly ITRs, rep and cap  
CC coding regions, are useful in production of recombinant viral vectors  
CC for gene delivery. These vectors can be used as gene therapy  
CC vectors, vaccine vectors or antisense delivery vectors. The AAV-1 does  
CC not induce the formation of neutralising antibodies specific to any  
CC serotype of AAV hence is useful for transforming host cells, and in the  
CC preparation of a medicament for the delivery of transgene to a host.  
CC The present sequence is a 5' ITR of AAV-1 DNA which is useful in the  
CC production of recombinant viral vector. The ITR forms a T-shaped hairpin  
CC structure.  
XX  
SQ Sequence 143 BP; 24 A; 50 C; 48 G; 21 T; 0 other;  
XX  
Query Match 100.0%; Score 143; DB 21; Length 143;  
Best Local Similarity 100.0%; Pred. No. 1e-28;  
Matches 143; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TTGCCCACTCCCTCTCTGCGCGCTCGCTCGCTGGGGCCCTGCGGACCAAGTCCGC 60  
DB 1 TTGCCCACTCCCTCTCTGCGCGCTCGCTCGCTGGGGCCCTGCGGACCAAGTCCGC 60  
QY 61 AGACGGCAGAGCTGTGCTGTGCGGCGCCACCGAGCGAGCGCGGAGAGGAGGTG 120  
DB 61 AGACGGCAGAGCTGTGCTGTGCGGCGCCACCGAGCGAGCGCGGAGAGGAGGTG 120  
QY 121 GGCAACTCCATCACTAGGGGTAA 143  
DB 121 GGCAACTCCATCACTAGGGGTAA 143  
RESULT 2  
AAD00772  
ID AAD00772 standard; DNA; 4718 BP.  
XX  
AC AAD00772;  
XX  
DT 08-SEP-2000 (first entry)  
XX  
DE Adeno-associated virus serotype 1 DNA.  
XX  
KW Adeno-associated virus serotype 1; AAV-1; rep protein; capsid protein;  
KW cap protein; recombinant viral vector; gene delivery; gene therapy;  
KW vaccine; transgene; ss.  
XX  
OS Adeno associated virus serotype 1.  
XX  
FH Key Location/Qualifiers  
FT repeat\_unit 1..143  
FT /\*tag= a  
FT /label= 5' ITR  
FT /note= "Inverted terminal repeat which is capable  
FT of forming T-shaped hairpin structure"  
FT protein\_bind 89..110

FT /\*tag= b  
FT /bound\_moiety= "Rep protein"  
FT misc\_feature 124..125  
FT /\*tag= c  
FT /note= "Terminal resolute site (TRS)"  
FT misc\_binding 219..226  
FT /\*tag= d  
FT /bound\_moiety= "USF"  
FT /note= "E box"  
FT promoter 236..299  
FT /\*tag= e  
FT /label= P5\_promoter  
FT protein\_bind 237..245  
FT /\*tag= f  
FT /bound\_moiety= "Y1 factor"  
FT TATA\_signal 270..275  
FT /\*tag= g  
FT /label= P5\_TATA-Box  
FT misc\_feature 299..306  
FT /\*tag= h  
FT /note= "Y1/p5 RNA"  
FT CDS 335..2206  
FT /\*tag= i  
FT /product= "Rep 78"  
FT /function= "regulates replication and integration  
FT of AAV DNA into host cell's chromosome"  
FT CDS 335..2272  
FT /\*tag= j  
FT /product= "Rep 68"  
FT /function= "regulates replication and integration  
FT of AAV DNA into host cell's chromosome"  
FT intron 1924..2220  
FT /\*tag= k  
FT /note= "This region interrupts the coding sequence  
FT of Rep 68 and Rep 40"  
FT TATA\_signal 857..862  
FT /\*tag= l  
FT /label= P19\_TATA\_Box  
FT misc\_feature 882..883  
FT /\*tag= m  
FT /note= "P19 RNA"  
FT CDS 1007..2206  
FT /\*tag= n  
FT /product= "Rep 52"  
FT /function= "regulates replication and integration  
FT of AAV DNA into host cell's chromosome"  
FT CDS 1007..2272  
FT /\*tag= o  
FT /product= "Rep 40"  
FT /function= "regulates replication and integration  
FT of AAV DNA into host cell's chromosome"  
FT TATA\_signal 1836..1841  
FT /\*tag= p  
FT /label= P40\_TATA-BOX  
FT misc\_feature 1875..1876  
FT /\*tag= q  
FT /note= "P40 RNA"  
FT CDS 2223..4433  
FT /\*tag= r  
FT /product= "VP1 protein"  
FT /note= "Capsid protein"  
FT CDS 2634..4433  
FT /\*tag= s  
FT /product= "VP2 protein"  
FT /note= "Capsid protein"  
FT CDS 2829..4433  
FT /\*tag= t  
FT /product= "VP3 protein"  
FT /note= "Capsid protein"  
FT polyA\_signal 4447..4452



```
FT      repeat_unit      4576..4718      /*tag= u
FT                                     /*tag= v
FT                                     /label= 3'_ITR
FT                                     /note= "Inverted terminal repeat which is capable
FT                                     of forming T-shaped hairpin structure"
XX
XX      WO200028061-A2.
XX
XX      18-MAY-2000.
XX
XX      02-NOV-1999;      99WO-US25694.
XX
XX      05-NOV-1998;      98US-0107114.
XX
XX      (UYPE-) UNIV PENNSYLVANIA.
XX
XX      Wilson JM, Xiao W;
XX
XX      WPI; 2000-376571/32.
XX
XX      P-PSDB; AAY71161, AAY71164, AAY71165, AAY71166, AAY71167, AAY71168,
XX      AAY71169.
XX
XX      Novel adeno-associated virus serotype 1 polynucleotide useful for
XX      preparation of medicament for delivery of a transgene to a host -
XX
XX      Claim 1; Fig 1; 108bp; English.
XX
XX      The present sequence is an adeno-associated virus serotype 1 (AAV-1)
XX      DNA characterised by two inverted terminal repeats (ITR) and open
XX      reading frames for rep and capsid (cap) proteins. The rep reading frame
XX      encodes four proteins, Rep 78, Rep 68, Rep 52 and Rep 40, while the cap
XX      reading frame encodes three structural proteins, VP1, VP2 and VP3.
XX      The AAV-1 sequence or its fragments particularly ITRs, rep and cap
XX      coding regions, are useful in production of recombinant viral vectors
XX      for gene delivery. These vectors can be used as gene therapy
XX      vectors, vaccine vectors or antisense delivery vectors. The AAV-1 does
XX      not induce the formation of neutralising antibodies specific to any
XX      serotype of AAV hence is useful for transforming host cells, and in the
XX      preparation of a medicament for the delivery of transgene to a host.
XX
XX      Sequence 4718 BP; 1121 A; 1393 C; 1273 G; 931 T; 0 other;
XX
XX      Query Match      100.0%; Score 143; DB 21; Length 4718;
XX      Best Local Similarity 100.0%; Pred. No. 1.2e-28;
XX      Matches 143; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX      1 TTGCCCACTCCCTCTCTGCGGCTCGCTCGGTGGGCGCTGCGGACCAAGGTCCGC 60
XX      1 TTGCCCACTCCCTCTCTGCGGCTCGCTCGGTGGGCGCTGCGGACCAAGGTCCGC 60
XX
XX      61 AGACGGCAGAGCTTGCTGTGCGCGGCCCAACGAGCGAGCGCGGAGAGAGGAGTG 120
XX      61 AGACGGCAGAGCTTGCTGTGCGCGGCCCAACGAGCGAGCGCGGAGAGAGGAGTG 120
XX
XX      121 GGCAACTCCATCACTAGGGTAA 143
XX      121 GGCAACTCCATCACTAGGGTAA 143
XX
XX      121 GGCAACTCCATCACTAGGGTAA 143
XX
XX      RESULT 3
XX      AAF23749/c
XX      ID AAF23749 standard; DNA; 4683 BP.
XX
XX      AAF23749;
XX
XX      28-MAR-2001 (first entry)
XX
XX      AAV6 DNA sequence.
XX
XX      AAV6; gene therapy; AAV viral vector; cystic fibrosis; cancer; AIDS;
XX      atherosclerosis; sickle cell anaemia; thalassaemia;
XX      blood clotting disorder; diabetes; ss.
```

```
XX      Adeno associated virus.
XX
XX      US6156303-A.
XX
XX      05-DEC-2000.
XX
XX      11-JUN-1997;      97US-0873168.
XX
XX      11-JUN-1997;      97US-0873168.
XX
XX      11-JUN-1997;      97US-0873168.
XX
XX      (UNIW ) UNIV WASHINGTON.
XX
XX      Russell DW, Rutledge EA;
XX
XX      WPI; 2001-060164/07.
XX
XX      Adeno-associated virus serotype 6 and viral vector derived from it for
XX      gene therapy of cystic fibrosis, cancer, acquired immunodeficiency
XX      syndrome, sickle cell anemia, thalassemia and diabetes -
XX
XX      Claim 1; Fig 1; 50pp; English.
XX
XX      The present invention relates to adeno-associated virus serotypes. The
XX      present sequence is the DNA sequence of one such serotype (AAV6). AAV6
XX      can be used to construct AAV viral vectors for use in gene therapy for a
XX      range of disorders: cystic fibrosis, cancer, AIDS, atherosclerosis,
XX      sickle cell anaemia, thalassaemia, blood clotting disorders and diabetes.
XX      The AAV viral vectors have increased transduction efficiency of a
XX      particular host cell as the AAV virion containing the AAV vector genome
XX      can be modified to express a capsid protein of an AAV serotype that
XX      transduces the selected host cell.
XX
XX      Sequence 4683 BP; 1114 A; 1363 C; 1277 G; 929 T; 0 other;
XX
XX      Query Match      97.1%; Score 138.8; DB 22; Length 4683;
XX      Best Local Similarity 98.6%; Pred. No. 1.5e-27;
XX      Matches 140; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX      1 TTGCCCACTCCCTCTCTGCGGCTCGCTCGGTGGGCGCTGCGGACCAAGGTCCGC 60
XX      4683 TTGCCCACTCCCTCTATGCGGCTCGCTCGGTGGGCGCTGCGGACCAAGGTCCGC 4624
XX
XX      61 AGACGGCAGAGCTTGCTGTGCGCGGCCCAACGAGCGAGCGCGGAGAGGAGTG 120
XX      4623 AGACGGCAGAGCTTGCTGTGCGCGGCCCAACGAGCGAGCGCGGAGAGGAGTG 4564
XX
XX      121 GGCAACTCCATCACTAGGGTAA 142
XX      4563 GGCAACTCCATCACTAGGGTAA 4542
XX
XX      RESULT 4
XX      AAD00772/c
XX      ID AAD00772 standard; DNA; 4718 BP.
XX
XX      AAD00772;
XX
XX      08-SEP-2000 (first entry)
XX
XX      Adeno-associated virus serotype 1 DNA.
XX
XX      Adeno-associated virus serotype 1; AAV-1; rep protein; capsid protein;
XX      cap protein; recombinant viral vector; gene delivery; gene therapy;
XX      vaccine; transgene; ss.
XX
XX      Adeno associated virus serotype 1.
XX
XX      Key      Location/Qualifiers
XX      repeat_unit      1..143
XX
XX      /*tag= a
XX      /label= 5'_ITR
XX      /note= "Inverted terminal repeat which is capable
```

FT	protein_bind	89..110	/of forming T-shaped hairpin structure"
FT		/tag= b	
FT	misc_feature	/bound_moiety= "Rep protein"	
FT		124..125	
FT		/tag= c	
FT	misc_binding	/note= "Terminal resolute site (TRS) "	
FT		219..226	
FT		/tag= d	
FT	promoter	/bound_moiety= "USF"	
FT		/note= "E box"	
FT		236..299	
FT		/tag= e	
FT	protein_bind	/label= P5_promoter	
FT		237..245	
FT		/tag= f	
FT	TATA_signal	/bound_moiety= "YY1 factor"	
FT		270..275	
FT		/tag= g	
FT		/label= P5_TATA-Box	
FT		299..306	
FT	misc_feature	/tag= h	
FT		/note= "YY1/p5 RNA"	
FT	CDS	335..2206	
FT		/tag= i	
FT		/product= "Rep 78"	
FT		/function= "regulates replication and integration of AAV DNA into host cell's chromosome"	
FT	CDS	335..2272	
FT		/tag= j	
FT		/product= "Rep 68"	
FT		/function= "regulates replication and integration of AAV DNA into host cell's chromosome"	
FT		/note= "The coding region is interrupted by intron"	
FT	intron	1924..2220	
FT		/tag= k	
FT		/note= "This region interrupts the coding sequence of Rep 68 and Rep 40"	
FT	TATA_signal	857..862	
FT		/tag= l	
FT		/label= P19_TATA_Box	
FT	misc_feature	882..883	
FT		/tag= m	
FT	CDS	/note= "P19 RNA"	
FT		1007..2206	
FT		/tag= n	
FT		/product= "Rep 52"	
FT		/function= "regulates replication and integration of AAV DNA into host cell's chromosome"	
FT	CDS	1007..2272	
FT		/tag= o	
FT		/product= "Rep 40"	
FT		/function= "regulates replication and integration of AAV DNA into host cell's chromosome"	
FT	TATA_signal	1836..1841	
FT		/tag= p	
FT		/label= P40_TATA-BOX	
FT	misc_feature	1875..1876	
FT		/tag= q	
FT		/note= "P40 RNA"	
FT	CDS	2223..4433	
FT		/tag= r	
FT		/product= "VP1 protein"	
FT		/note= "Capsid protein"	
FT	CDS	2634..4433	
FT		/tag= s	
FT		/product= "VP2 protein"	
FT		/note= "Capsid protein"	
FT		/partial	
FT	CDS	2829..4433	
FT		/tag= t	
FT		/product= "VP3 protein"	

FT		/note= "Capsid protein"	
FT	polyA_signal	4447..4452	
FT		/*tag= u	
FT	repeat_unit	4576..4718	
FT		/*tag= v	
FT		/label= 3' _ITR	
FT		/note= "Inverted terminal repeat which is capable of forming T-shaped hairpin structure"	
PN	WO200028061-A2.		
XX			
PD	18-MAY-2000.		
XX			
PF	02-NOV-1999;	99WO-US25694.	
XX			
PR	05-NOV-1998;	98US-0107114.	
XX			
PA	(UYPE-) UNIV PENNSYLVANIA.		
PI	Wilson JM, Xiao W;		
XX			
XX	WPI; 2000-376571/32.		
DR	P-PSDB; AAY71161, AAY71164,	AAY71165, AAY71166, AAY71167, AAY71168,	
DR	AAY71169.		
XX			
PT	Novel adeno-associated virus serotype 1 polynucleotide useful for		
PT	preparation of medicament for delivery of a transgene to a host		
PS	Claim 1; Fig 1; 108pp; English.		
XX			
XX	The present sequence is an adeno-associated virus serotype 1 (AAV-1)		
CC	DNA characterised by two inverted terminal repeats (ITR) and open		
CC	reading frames for rep and capsid (cap) proteins. The rep reading frame		
CC	encodes four proteins, Rep 78, Rep 68, Rep 52 and Rep 40, while the cap		
CC	reading frame encodes three structural proteins, VP1, VP2 and VP3.		
CC	The AAV-1 sequence or its fragments particularly ITRs, rep and cap		
CC	coding regions, are useful in production of recombinant viral vectors		
CC	for gene delivery. These vectors can be used as gene therapy		
CC	vectors, vaccine vectors or antisense delivery vectors. The AAV-1 does		
CC	not induce the formation of neutralising antibodies specific to any		
CC	serotype of AAV hence is useful for transforming host cells, and in the		
CC	preparation of a medicament for the delivery of transgene to a host.		
XX			
SQ	Sequence 4718 BP; 1121 A; 1393 C; 1273 G; 931 T; 0 other;		
	Query Match	83.2%;	Score 119; DB 21; Length 4718;
	Best Local Similarity	89.5%;	Pred. No. 2.4e-22;
	Matches 128; Conservative	0;	Mismatches 15; Indels 0; Gaps 0;
QY	1 TTGCCACTCCCTCTCTCTGCGCGCTCGCTCGCTGGGGCTGCGGACCAAGGTC	60	
Db	4718 TTGCCACTCCCTCTCTCTGCGCGCTCGCTCGCTGGGGCGCGAGAGAGCTCTGC	4659	
QY	61 AGACGGCAGAGCTCTGCTCTGCGCGGCCCAACGAGCGAGCGCGCAGAGGAGTG	120	
Db	4658 CGTCTGCGGACCTTTGGTCCGAGAGGCCCAACGAGCGAGCGCGCAGAGGAGTG	4599	
QY	121 GGCACTCCATCACTAGGGGTAA	143	
Db	4598 GGCACTCCATCACTAGGGGTAA	4576	
	RESULT 5		
AAF23748/C	AAF23748 standard; DNA; 4722 BP.		
XX	AAF23748;		
AC			
XX			
DT	28-MAR-2001 (first entry)		
XX			
DE	AAV3B DNA sequence.		
XX			
KM	AAV3B; gene therapy; AAV viral vector; cystic fibrosis; cancer; AIDS;		

KW atherosclerosis; sickle cell anaemia; thalassaemia;  
KM blood clotting disorder; diabetes; ss.  
XX  
OS Adeno associated virus.  
XX  
PN US6156303-A.  
XX  
PD 05-DEC-2000.  
XX  
PF 11-JUN-1997; 97US-0873168.  
XX  
PR 11-JUN-1997; 97US-0873168.  
XX  
PA (UNIW ) UNIV WASHINGTON.  
XX  
PI Russell DW, Rutledge EA;  
XX  
DR WPI; 2001-060164/07.  
XX  
PT Adeno-associated virus serotype 6 and viral vector derived from it for  
PT gene therapy of cystic fibrosis, cancer, acquired immunodeficiency  
PT syndrome, sickle cell anemia, thalassemia and diabetes  
XX  
PS Example 2; Fig 1; 50pp; English.  
XX  
CC The present invention relates to adeno-associated virus serotypes. The  
CC present sequence is the DNA sequence of one such serotype (AAV3B). AAV3B  
CC can be used to construct AAV viral vectors for use in gene therapy for a  
CC range of disorders: cystic fibrosis, cancer, AIDS, atherosclerosis,  
CC sickle cell anaemia, thalassaemia, blood clotting disorders and diabetes.  
CC The AAV viral vectors have increased transduction efficiency of a  
CC particular host cell as the AAV virion containing the AAV vector genome  
CC can be modified to express a capsid protein of an AAV serotype that  
CC transduces the selected host cell.  
XX  
SQ Sequence 4722 BP; 1204 A; 1297 C; 1243 G; 978 T; 0 other;  
  
Query Match 79.2%; Score 113.2; DB 22; Length 4722;  
Best Local Similarity 87.3%; Pred. No. 7.8e-21;  
Matches 124; Conservative 0; Mismatches 18; Indels 0; Gaps 0;  
  
QY 1 TTGCCACTCCCTCTCTGCGCGCTCGCTCGGTGGGGCTGCGGACCAAGTCCGC 60  
DB 4722 TTGGCCACTCCCTCTATGCGCACTCGCTCGGTGGGGCTGCGGACCAAGTCCGC 4663  
  
QY 61 AGACGGCAGAGCTCTGCTCTGCGCGCCGCCACGAGCGAGCGCGCAGAGAGGAGTG 120  
DB 4662 AGACGGCAGAGCTCTTGACAGTCCGCCGCCACGAGCGAGTGCCTAGAGGAGTG 4603  
  
QY 121 GGCAACTCCATCACTAGGGGTA 142  
DB 4602 GCCAACTCCATCACTAGAGGTA 4581  
  
RESULT 6  
AAF23748  
ID AAF23748 standard; DNA; 4722 BP.  
XX  
AC AAF23748;  
XX  
DT 28-MAR-2001 (first entry)  
XX  
DE AAV3B DNA sequence.  
XX  
KM AAV3B; gene therapy; AAV viral vector; cystic fibrosis; cancer; AIDS;  
KM atherosclerosis; sickle cell anaemia; thalassaemia;  
KM blood clotting disorder; diabetes; ss.  
XX  
OS Adeno associated virus.  
XX  
PN US6156303-A.  
XX  
PD 05-DEC-2000.

XX  
PF 11-JUN-1997; 97US-0873168.  
XX  
PR 11-JUN-1997; 97US-0873168.  
XX  
PA (UNIW ) UNIV WASHINGTON.  
XX  
PI Russell DW, Rutledge EA;  
XX  
DR WPI; 2001-060164/07.  
XX  
PT Adeno-associated virus serotype 6 and viral vector derived from it for  
PT gene therapy of cystic fibrosis, cancer, acquired immunodeficiency  
PT syndrome, sickle cell anemia, thalassemia and diabetes  
XX  
PS Example 2; Fig 1; 50pp; English.  
XX  
CC The present invention relates to adeno-associated virus serotypes. The  
CC present sequence is the DNA sequence of one such serotype (AAV3B). AAV3B  
CC can be used to construct AAV viral vectors for use in gene therapy for a  
CC range of disorders: cystic fibrosis, cancer, AIDS, atherosclerosis,  
CC sickle cell anaemia, thalassaemia, blood clotting disorders and diabetes.  
CC The AAV viral vectors have increased transduction efficiency of a  
CC particular host cell as the AAV virion containing the AAV vector genome  
CC can be modified to express a capsid protein of an AAV serotype that  
CC transduces the selected host cell.  
XX  
SQ Sequence 4722 BP; 1204 A; 1297 C; 1243 G; 978 T; 0 other;  
  
Query Match 78.5%; Score 112.2; DB 22; Length 4722;  
Best Local Similarity 87.2%; Pred. No. 1.4e-20;  
Matches 123; Conservative 0; Mismatches 18; Indels 0; Gaps 0;  
  
QY 2 TGCCCACTCCCTCTCTGCGCGCTCGCTCGGTGGGGCTGCGGACCAAGTCCGCA 61  
DB 1 TGCCCACTCCCTCTATGCGCACTCGCTCGGTGGGGCTGCGGACCAAGTCCGCA 60  
  
QY 62 GACGGCAGAGCTCTGCTCTGCGCGCCGCCACGAGCGAGCGCGCAGAGAGGAGTG 121  
DB 61 GACGGCAGAGCTCTTGACAGTCCGCCGCCACGAGCGAGTGCCTAGAGGAGTG 120  
  
QY 122 GCAACTCCATCACTAGGGGTA 142  
DB 121 CCAACTCCATCACTAGAGGTA 141  
  
RESULT 7  
AAQ41448  
ID AAQ41448 standard; DNA; 145 BP.  
XX  
AC AAQ41448;  
XX  
DT 27-AUG-1993 (first entry)  
XX  
DE AAV2 inverted terminal repeat.  
XX  
KM Adeno-associated virus 2; ITR; site-specific integration; vector;  
KM cell-specific; gene therapy; haemoglobinopathies; thalassaemia;  
KM diabetes; sickle cell anaemia; cancer; parvovirus; B19; ss.  
XX  
OS Adeno-associated virus 2.  
XX  
FH Key  
FH -repeat\_region 1..125  
FT /\*tag= a  
FT /note= "inverted terminal repeat forming  
FT misc\_feature 42..83 palindromic hairpin"  
FT /\*tag= b  
FT /note= "Flip orientation"  
XX  
PN WO9309239-A.  
XX



PD 13-MAY-1993.  
XX  
PF 06-NOV-1992; 92WO-US09769.  
XX  
PR 08-NOV-1991; 91US-0789917.  
XX  
PA (RESE ) RESEARCH CORP TECHNOLOGIES INC.  
XX  
PI Srivastava A;  
XX  
DR WPI, 1993-167704/20.  
XX  
PT Adeno-associated virus-2 basal vectors - for gene therapy and  
PT treatment of haemoglobinopathies and cancer etc. - has cassette  
PT contg. a promoter capable of cell-specific expression, between 2  
PT inverted terminal repeats of the adeno-associated virus 2  
XX  
PS Disclosure; Fig 1; 68bp; English.  
XX  
CC The sequence is one of two inverted terminal repeat sequences, one  
CC from parvovirus B19 and the other from adeno-associated virus 2  
CC (AAV2), used in construction of an expression vector for site  
CC specific integration and cell specific gene expression. The  
CC vector comprises at least one cassette contg. a promoter capable  
CC of effecting cell-specific expression, operably linked to a hetero-  
CC logous gene, and the cassette residing between the inverted terminal  
CC repeats. The vector is safe for use in gene therapy, partic. in  
CC treatment of haemoglobinopathies and a variety of diseases, e.g.  
CC thalassemia, diabetes, sickle cell anaemia, and cancer.  
CC See also AAQ41449.  
XX  
SQ Sequence 145 BP; 21 A; 52 C; 49 G; 23 T; 0 other;  
XX  
Query Match 75.1%; Score 107.4; DB 14; Length 145;  
Best Local Similarity 85.1%; Pred. No. 2.2e-19;  
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;  
QY 1 TTGCCCACTCCCTCTGTGCGCGCTCGCTCGCTCGGTGGGCGCTGCGGACCAAGTCCGC 60  
DB 1 TTGGCCACTCCCTCTGTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGAACCAAGTCCGC 60  
QY 61 AGACGGCAGAGCTGTGCTGTGCGCGCGCGCGCGCGCGAGCGAGCGGCGAGAGGAGTG 120  
DB 61 CGACGCCCGGGCTTTGCCCGGGCGCGCTCACTGAGCGAGCGAGCGGCGAGAGGAGTG 120  
QY 121 GGCAACTGCATCACTAGGGGT 141  
DB 121 GCCAACTCCATCACTAGGGGT 141  
RESULT 8  
AAT03385  
ID AAT03385 standard; cDNA; 145 BP.  
XX  
AC AAT03385;  
XX  
DT 18-APR-1996 (first entry)  
XX  
DE Strict inverted terminal repeat from AAV-2, used in pITRFL.  
XX  
KW inverted terminal repeat; ITR; adenovirus; adeno-associated virus;  
KW AAV-2; replication defective; integration; gene therapy; ds.  
XX  
OS Adeno associated virus 2.  
XX  
PN WO9523867-A1.  
XX  
PD 08-SEP-1995.  
XX  
PF 28-FEB-1995; 95WO-FR00233.  
XX  
PR 03-MAR-1994; 94FR-0002445.  
XX

PA (RHON ) RHONE POULENC RORER SA.  
XX  
PI Deneffe P, Latta M, Perricaudet M, Vigne E;  
XX  
DR WPI, 1995-320581/41.  
XX  
PT Recombinant defective adenovirus contg. integratable expression  
PT cassette - for use in gene therapy to express protein, antigen or  
PT anti:sense nucleic acid, also for prodn. of recombinant  
PT adeno-associated viruses  
XX  
PS Example 3; Page 26; 50pp; French.  
XX  
CC Recombinant, non-pathogenic adenovirus which are able to integrate  
CC stably into a host genome are claimed. The viruses pref. contain at  
CC least one inverted terminal repeat (ITR) sequence and in particular  
CC two ITRs flank a heterologous DNA insert. The present sequence is  
CC that of the strict ITR from adeno associated virus AAV-2 (i.e. the  
CC ITR sequence without any deletions or additions). In the construct  
CC pITRFL, the beta-galactosidase marker gene is flanked by two strict  
CC AAV-2 ITRs. The defective viruses are useful for stably introducing  
CC large fragments of heterologous DNA making them suitable for gene  
CC therapy.  
XX  
SQ Sequence 145 BP; 21 A; 52 C; 49 G; 23 T; 0 other;  
XX  
Query Match 75.1%; Score 107.4; DB 16; Length 145;  
Best Local Similarity 85.1%; Pred. No. 2.2e-19;  
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;  
QY 1 TTGCCCACTCCCTCTGTGCGCGCTCGCTCGCTCGGTGGGCGCTGCGGACCAAGTCCGC 60  
DB 1 TTGGCCACTCCCTCTGTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGAACCAAGTCCGC 60  
QY 61 AGACGGCAGAGCTGTGCTGTGCGCGCGCGCGCGCGAGCGAGCGGCGAGAGGAGTG 120  
DB 61 CGACGCCCGGGCTTTGCCCGGGCGCGCTCACTGAGCGAGCGAGCGGCGAGAGGAGTG 120  
QY 121 GGCAACTGCATCACTAGGGGT 141  
DB 121 GCCAACTCCATCACTAGGGGT 141  
RESULT 9  
AAAX34295  
ID AAAX34295 standard; DNA; 145 BP.  
XX  
AC AAAX34295;  
XX  
DT 16-JUL-1999 (first entry)  
XX  
DE Adeno-associated virus inverted terminal repeat sequence.  
XX  
KW Recombinant; chimeric; parvovirus; adeno-associated virus; AAV; vector;  
KW promoter; rep; cap; inverted terminal repeat; ITR; erythroid cell;  
KW integration; gene expression; bone marrow; peripheral blood cell;  
KW endothelial cell; myocardial cell; ss.  
XX  
OS Adeno associated virus.  
XX  
PN WO9918227-A1.  
XX  
PD 15-APR-1999.  
XX  
PF 08-OCT-1998; 98WO-US21202.  
XX  
PR 08-OCT-1997; 97US-0061364.  
XX  
PA (ADRE-) ADVANCED RES & TECHNOLOGY INST.  
XX  
PI Ponnazhagan S, Srivastava A;  
XX  
DR WPI, 1999-264033/22.  
XX

XX	New recombinant adeno-associated vectors
PT	Claim 2; Page 69; 76pp; English.
XX	The invention relates to new recombinant chimeric parvovirus-adeno-
PS	associated virus (AAV) vectors comprising a promoter e.g. an AAV
XX	promoter (AAK34296) and a selected DNA sequence, especially an AAV rep
CC	gene and a parvovirus B19 cap gene, located between 2 AAV inverted
CC	terminal repeats (ITR) such as the ITR sequence shown here. The system
CC	can specifically target primitive progenitor and differentiated cells of
CC	the erythroid lineage, and can achieve stable integration and expression
CC	of transduced genes. The vectors can be used for the in vitro or in vivo
CC	delivery of genes to cells such as bone marrow cells, peripheral blood
CC	cells, endothelial cells and myocardial cells.
XX	
SQ	Sequence 145 BP; 21 A; 52 C; 49 G; 23 T; 0 other;
	Query Match                      75.1%; Score 107.4; DB 20; Length 145;
	Best Local Similarity        85.1%; Pred. No. 2.2e-19;
	Matches 120; Conservative     0; Mismatches 21; Indels 0; Gaps 0;
OY	1 TTGCCACTCCTCTCTGCGGCTGCTGCTCGTGGGGCTGCGGACCAAGGTCCGC 60
Db	1 TTGGCCACTCCCTCTCTGCGGCTGCTGCTCGTCACTGAGGCCGGGCGAACCAAGTCCGC 60
OY	61 AGACGGCAGAGCTCTGCTCTGCGGCGCCACCACCGCAGCGAGCGCGCAGAGAGGAGTG 120
Db	61 CGACGCCCGGGCTTTGGCCCGGGCGGCTCACTGAGCGAGCGAGCGCGCAGAGAGGAGTG 120
OY	121 GGCACTCCATCATCTAGGGGT 141
Db	121 GCCAATCCATCATCTAGGGGT 141
	RESULT 10
ID	AAQ66769 standard; DNA; 165 BP.
XX	AAQ66769;
AC	
XX	20-JAN-1995 (first entry)
DT	
XX	Double-D ITR terminal repeat.
DE	
XX	Double-D ITR; inverted terminal repeat; D-sequence; AAV;
KW	adeno-associated virus; capsid; encapsidation; gene therapy; vector;
KM	ds.
XO	
OS	Synthetic.
PN	WO9413788-A.
PD	23-JUN-1994.
PF	03-DEC-1993; 93WO-US11728.
PR	04-DEC-1992; 92US-0989841.
PA	(UYPI-) UNIV PITTSBURGH.
PI	
XX	Samulski RJ, Xiao X;
DR	WPI, 1994-217868/26.
XX	
PT	Adenovirus associated viral inverted terminal repeat - for use in
PT	a recombinant viral vector system for treatment of genetic
PT	diseases
XX	
PS	Disclosure; Page 25; 44pp; English.
XX	
XX	The 20-bp D-sequence given in AAQ66773 is present in the inverted
CC	terminal repeat (ITR) sequence of AAV and is required for viral
CC	

Query	Match	Score	DB	Length
Best Local Similarity	85.1%	Pred. No. 2.2e-19		
Matches	120	Conservative	0	Mismatches 21; Indels 0; Gaps 0
Db	1	TTGCCACTCCCTCTCTGCGCGCTCGCTCGCTCGGTGGGCGCTCGGACCAAGGTCGC	60	
	21	TTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCGCGGACCAAGGTCGCC	80	
Qy	61	AGACGGCAGAGCTCTGCTCTGCGCGCGCGCGCGCGAGCGAGCGCGCAGAGAGGAGTG	120	
Db	81	CGACGCCCGCGGCTTGGCCGGCGCGCGCTCACTGAGCGAGCGAGCGCGCAGAGAGGAGTG	140	
Qy	121	GGCACTCCATCCTAGGGGT	141	
Db	141	GCCCACTCCATCCTAGGGGT	161	
RESULT 11				
AAT49462	AAT49462 standard; DNA; 165 BP.			
AC	AAT49462;			
XX				
DT	10-SEP-1997 (first entry)			
XX				
DE	Adeno associated virus inverted terminal repeat with double D region.			
XX				
KM	Viral replication; REP protein; inverted terminal repeat; ITR;			
KM	adeno-associated virus; AAV; vector; double-D; cis-acting;			
KW	lytic life cycle; gene therapy; ss.			
XX				
OS	Adeno associated virus.			
XX				
PH	Key			
FT	repeat_unit			
FT	Location/Qualifiers			
FT	1..145			
FT	/*tag= a			
FT	/rpt_type= INVERTED			
FT	/standard_name= ITR			
FT	/note= "This naturally occurring 145 bp ITR is			
FT	located at both ends of the AAV genome"			
FT	1..20			
FT	/*tag= b			
FT	/rpt_type= INVERTED			
FT	/label= D_repeat			
FT	21..62			
FT	/*tag= c			
FT	/rpt_type= INVERTED			
FT	/label= A_repeat			
FT	/note= "Forms the stem of a T-shaped structure when			
FT	base paired with repeat A"			
FT	63..70			
FT	/*tag= d			
FT	/rpt_type= INVERTED			
FT	/label= C_repeat			
FT	74..81			
FT	/*tag= e			
FT	/rpt_type= INVERTED			
FT	/label= C_repeat			
FT	85..92			
FT	/*tag= f			
FT	/rpt_type= INVERTED			
FT	/label= B_repeat			
FT	96..103			
FT	repeat_unit			

FT FT /\*tag= 9  
FT /rpt\_type= INVERTED  
FT /label= B'\_repeat  
FT repeat\_unit 104..145  
FT /\*tag= h  
FT /rpt\_type= INVERTED  
FT /label= A'\_repeat  
FT /note= "Forms the stem of a T-shaped structure when  
base paired with A' repeat"  
FT repeat\_unit 146..165  
FT /\*tag= 1  
FT /label= D'  
FT /note= "Additional D' sequence"  
FT stem\_loop 61..82  
FT /\*tag= j  
FT /note= "Part of T-shaped structure, formed by base  
pairing between the C' and C repeats"  
FT stem\_loop 84..104  
FT /\*tag= k  
FT /note= "Part of T-shaped structure, formed by base  
pairing between the B' and B repeats"  
FT misc\_structure 21..145  
FT /\*tag= 1  
FT /label= T-shaped  
FT /note= "The naturally occurring 145 bp ITR sequence  
can form a T-shaped structure for DNA  
replication when single-stranded"  
FT repeat\_region 1..165  
FT /\*tag= m  
FT /label= Double-D  
FT /note= "This 165 bp sequence, resulting from  
addition of a second D repeat to the AAV  
ITR, has not been identified in any  
naturally occurring virus"  
XX PN WO9636364-A1.  
XX PD 21-NOV-1996.  
XX PF 14-MAY-1996; 96WO-US06786.  
XX PR 15-MAY-1995; 95US-0440738.  
XX PA (SAMU/) SAMULSKI R J.  
XX PA (XIAO/) XIAO X.  
XX PI Samulski RJ, Xiao X;  
XX DR WPI; 1997-042643/04.  
XX XX  
XX PT Double-D sequence directs adeno-associated virus integration into  
PT host genome - used in gene therapy, maintains full length coding  
PT sequence of therapeutic gene  
XX PS Claim 1; Fig 9; 55pp; English.  
XX XX  
CC The adeno associated virus (AAV) has a 145 bp inverted terminal  
CC repeat (ITR) located at each end of its genome. In addition to  
CC being able to base pair with each other, the ITRs can also  
CC individually fold back on themselves through the base pairing of A,  
CC A', B, B' and C, C' sequences to form a T-shaped structure for DNA  
CC replication (see features table). It has been found that viral  
CC mutants with deleted D sequences are unable to replicate their DNA.  
CC Addition of a second D repeat (D') to the naturally occurring ITR  
CC resulted in a sequence, designated double-D, which was sufficient  
CC to carry out the functions normally required of two wild-type ITRs  
CC during a lytic AAV viral infection, i.e. it is capable of directing  
CC replication and assembly into AAV, and/or the integration into the  
CC host genome, of recombinant DNA containing the nucleic acid molecule.  
CC Vectors and viral particles containing the double-D sequence are  
CC useful in gene therapy. Replication and integration into the host  
CC genome is completely effected through the double-D sequences, ensuring  
CC that the heterologous gene sequences remain intact.

XX SQ Sequence 165 BP; 27 A; 56 C; 56 G; 26 T; 0 other;  
Query Match 75.1%; Score 107.4; DB 18; Length 165;  
Best Local Similarity 85.1%; Pred. No. 2.2e-19;  
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;  
QY 1 TTGCCCACTCCCTCTCTGCGCGCTGCTGCTCGGTGGGGCTGGCGCAAGTCCGC 60  
DB 21 TTGGCCACTCCCTCTCTGCGCGCTGCTGCTGCTCACTGAGGCCCGGCGCAAGTCCGC 80  
QY 61 AGACGGCAGAGCTCTGCTCTGCGCGCCCAACCGAGCGAGCGCGCGAGAGGAGTG 120  
DB 81 CGACGCCCGGCTTTGCCCCGGCGGCTCACTGAGCGAGCGAGCGCGAGAGGAGTG 140  
QY 121 GGCAACTCCATCACTAGGGGT 141  
DB 141 GCCAAGTCCATCACTAGGGGT 161  
RESULT 12  
AAD03535  
ID AAD03535 standard; DNA; 165 BP.  
XX AC AAD03535;  
XX DT 19-JUN-2001 (first entry)  
XX DE Inverted terminal repeat double DB DNA sequence.  
XX XX  
KW Recombinant adeno-associated virus; rAAV; circular intermediate; ITR;  
KW Inverted terminal repeat; haemostatic; antisickling; neuroprotective;  
KW antianaemic; nootropic; blood disorder; sickle cell anaemia;  
KW thalassaemia; neurological disorder; haemophilia; Alzheimer's disease;  
KW muscle disorder; Parkinson's disease; gene delivery; erythropoietin;  
KW epo; CFTR; cystic fibrosis transmembrane conductance receptor;  
KW tyrosine hydroxylase; glucocerebrosidase gene; Gaucher's disease;  
KW gene therapy; Double DD; ds.  
XX OS Unidentified.  
XX OS  
XX PN WO200125465-A1.  
XX PD 12-APR-2001.  
XX PF 06-OCT-2000; 2000WO-US27863.  
XX PR 07-OCT-1999; 99US-0158209.  
XX PA (IOWA ) UNIV IOWA RES FOUND.  
XX PA (ENGELHARDT J F.  
XX PA (DONG/) DONGSHENG D.  
XX PA (ZIYI/) ZIYING Y.  
XX PI Engelhardt JF, Dongsheng D, Ziyang Y;  
XX XX  
XX DR WPI; 2001-266321/27.  
XX XX  
XX PT Composition for transferring recombinant DNAs and to express a  
XX PT polypeptide in a host cell, comprises two recombinant adeno-associated  
XX PT viruses -  
XX PS Disclosure; Page 139; 144pp; English.  
XX XX  
CC The patent discloses a composition comprising at least two recombinant  
CC adeno-associated viruses (rAAVs). Each rAAV contains a recombinant DNA  
CC molecule comprising three DNA segments linked together. The first and  
CC the third DNA segments comprise 5' and 3' inverted terminal repeats  
CC (ITRs) respectively from the circular intermediate of AAV. The second  
CC DNA segment in each virus is different and does not comprise AAV  
CC sequence. It preferably comprises sequences encoding a therapeutically  
CC effective polypeptide such as the cystic fibrosis transmembrane  
CC conductance receptor gene (CFTR), the erythropoietin (epo) gene, the



CC tyrosine hydroxylase gene (Parkinson's syndrome) or the glucocere-  
CC brosidase gene (Gaucher's disease). The circular intermediate of AAV  
CC imparts increased episomal stability and persistence of the vector in  
CC the host cell. Compositions comprising rAAV sequences are useful for  
CC transferring recombinant DNAs to a host cell and express a polypeptide  
CC in a host cell. The recombinant vector is useful in medical therapy,  
CC which includes treatment or prophylaxis of blood disorders (e.g. sickle  
CC cell anaemia, thalassemia, haemophilia), neurological disorders, such  
CC as Alzheimer's disease, Parkinson's disease, muscle disorders involving  
CC skeletal, cardiac or smooth muscle. AAV vector is used as a delivery  
CC vehicle for gene therapy.  
CC The present sequence is the ITR DD DNA sequence referred to as  
CC "double sequence". This sequence is disclosed in U.S. Patent No.  
CC 5,478,745. This sequence is not used in the present invention.  
XX  
SQ Sequence 165 BP; 27 A; 56 C; 56 G; 26 T; 0 other;

Query Match 75.1%; Score 107.4; DB 22; Length 165;  
Best Local Similarity 85.1%; Pred. No. 2.2e-19;  
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

OY 1 TTGCCCACTCCCTCTCTGCGCGCTCGCTCGCTGCGCGCTGCGGACCAAGTCCGC 60  
|||  
DB 21 TTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGCGCGCGGCAACCAAGTCCGC 80  
OY 61 AGACGGCAGAGCTCTGCTCTGCGCGCGCGCGCGCGAGCGAGCGCGGAGAGAGAGTGTG 120  
|||  
DB 81 CGACGCCCGGGCTTTGCGCGCGCGCTCTCACTGAGCGAGCGCGCGGAGAGAGAGAGTGTG 140  
OY 121 GGCACTCCATCACTAGGGGT 141  
|||  
DB 141 GCCAATCCATCACTAGGGGT 161

RESULT 13

ID AAZ47166 standard; DNA; 272 BP.

XX AAZ47166;

DT 28-MAR-2000 (first entry)

XX Head-to-tail ITR sequence from recombinant AAV vector p1202.

XX Antisickling; hemostatic; neuroprotective; antiparkinsonian; ITR;  
KW gene therapy; inverted terminal repeat; vector; episomal stability;  
KW blood; neurological; muscle; disorder; ds.

XX Adeno associated virus.

XX WO9960146-A1.

XX 25-NOV-1999.

XX 20-MAY-1999; 99WO-US11197.

XX 20-MAY-1998; 98US-0086166.

XX 25-MAR-1999; 99US-0276625.

XX (IOWA ) UNIV IOWA RES FOUND.

XX (ENGE/) ENGELHARDT J F.

XX (DUAN/) DUAN D.

XX Engelhardt JF, Duan D;

XX WPI; 2000-062462/05.

PT Recombinant adeno-associated virus vector useful for gene therapy  
PT against disorders related to blood, neurological and muscular systems

XX Disclosure; Fig 10C; 121pp; English.

CC This sequence corresponds to a head-to-tail inverted terminal repeat  
CC (ITR) sequence from the recombinant adeno-associated virus (rAAV)  
CC vector p1202. The invention relates to a DNA molecule comprising a DNA  
CC segment (or biologically active subunit or variant) of a circular  
CC intermediate of an AAV that confers increased episomal stability,  
CC persistence or abundance in a host cell. Compositions containing a  
CC vector with therapeutic gene and delivery vehicle or containing two  
CC vectors expressing a full length polypeptide coordinately, are useful  
CC for manufacturing a medicament for treating pathological conditions or  
CC symptoms in a mammal. The vector is useful for therapeutic or  
CC prophylactic treatments of blood disorders (e.g. sickle cell anemia,  
CC thalassemias, hemophilias and Fanconi's anemias), neurological disorders  
CC (e.g Alzheimer's disease, Parkinson's disease) and muscle disorders.  
XX  
SQ Sequence 272 BP; 58 A; 78 C; 78 G; 58 T; 0 other;

Query Match 75.1%; Score 107.4; DB 21; Length 272;  
Best Local Similarity 85.1%; Pred. No. 2.3e-19;  
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

OY 1 TTGCCCACTCCCTCTCTGCGCGCTCGCTCGCTGCGCGCTGCGGACCAAGTCCGC 60  
|||  
DB 69 TTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGCGCGCGGCAACCAAGTCCGC 128  
OY 61 AGACGGCAGAGCTCTGCTCTGCGCGCGCGCGCGCGAGCGAGCGCGGAGAGAGAGTGTG 120  
|||  
DB 129 CGACGCCCGGGCTTTGCTCGCGCGCGCTCTCACTGAGCGAGCGCGGAGAGAGAGTGTG 188  
OY 121 GGCACTCCATCACTAGGGGT 141  
|||  
DB 189 GCCAATCCATCACTAGGGGT 209

RESULT 14

ID AAD03534 standard; DNA; 272 BP.

XX AAD03534;

DT 19-JUN-2001 (first entry)

XX AAV circular intermediate DNA segment comprising ITRs from clone p1202.

XX Recombinant adeno-associated virus; rAAV; circular intermediate; ITR;  
KW inverted terminal repeat; haemostatic; antisickling; neuroprotective;  
KW antianaemic; nootropic; blood disorder; sickle cell anaemia;  
KW thalassemia; neurological disorder; haemophilia; Alzheimer's disease;  
KW muscle disorder; Parkinson's disease; gene delivery; erythropoietin;  
KW epo; CFTR; cystic fibrosis transmembrane conductance receptor;  
KW tyrosine hydroxylase; glucocerebrosidase gene; Gaucher's disease;  
KW gene therapy; clone p1202; ds.

XX Adeno associated virus.

XX WO200125465-A1.

XX 12-APR-2001.

XX 06-OCT-2000; 2000WO-US27863.

XX 07-OCT-1999; 99US-0158209.

XX (IOWA ) UNIV IOWA RES FOUND.

XX (ENGE/) ENGELHARDT J F.

XX (DONG/) DONGSHENG D.

XX (ZIYI/) ZIYING Y.

XX Engelhardt JF, Dongsheng D, Ziyang Y;

PT Composition for transferring recombinant DNAs and to express a  
PT polypeptide in a host cell, comprises two recombinant adeno-associated

PT viruses -  
XX  
PS Disclosure; Fig 10C; 144pp; English.  
XX  
CC The patent discloses a composition comprising at least two recombinant  
CC adeno-associated viruses (AAVs). Each AAV contains a recombinant DNA  
CC molecule comprising three DNA segments linked together. The first and  
CC the third DNA segments comprise 5' and 3' inverted terminal repeats  
CC (ITRs) respectively from the circular intermediate of AAV. The second  
CC DNA segment in each virus is different and does not comprise AAV  
CC sequence. It preferably comprises sequences encoding a therapeutically  
CC effective polypeptide such as the cystic fibrosis transmembrane  
CC conductance receptor gene (CFTR), the erythropoietin (epo) gene, the  
CC tyrosine hydroxylase gene (Parkinson's syndrome) or the glucocere-  
CC brosidase gene (Gaucher's disease). The circular intermediate of AAV  
CC imparts increased episomal stability and persistence of the vector in  
CC the host cell. Compositions comprising AAV sequences are useful for  
CC transferring recombinant DNAs to a host cell and express a polypeptide  
CC in a host cell. The recombinant vector is useful in medical therapy,  
CC which includes treatment or prophylaxis of blood disorders (e.g. sickle  
CC cell anaemia, thalassemia, haemophilia), neurological disorders, such  
CC as Alzheimer's disease, Parkinson's disease, muscle disorders involving  
CC skeletal, cardiac or smooth muscle. AAV vector is used as a delivery  
CC vehicle for gene therapy.  
CC The present sequence is circular intermediate DNA segment comprising  
CC ITRs from Adeno associated virus (AAV) clone p1202.  
XX  
SQ Sequence 272 BP; 58 A; 78 C; 78 G; 58 T; 0 other;

Query Match 75.1%; Score 107.4; DB 22; Length 272;  
Best Local Similarity 85.1%; Pred. No. 2.3e-19;  
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 1 TTGCCACTCCCTCTCTGCGGCTCGCTCGCTGGGGGCTGCGACCAAGTCCGC 60  
DB 69 TTGGCCACTCCCTCTCTGCGGCTCGCTCGCTGAGGCCGGGCGACCAAGTCCGC 128  
QY 61 AGACGGCAGAGCTCTGCTCTGCGGCGCCACCGAGCGAGCGCGCAGAGGAGTG 120  
DB 129 CGACGCGCGGGCTTGTGCGCCCGGCTCAGTGAGCGAGCGCGCAGAGGAGTG 188  
QY 121 GGCACTCCATCACTAGGGGT 141  
DB 189 GCCAACTCCATCACTAGGGGT 209

RESULT 15  
AAD37254  
ID AAD37254 standard; DNA; 955 BP.  
XX  
AC AAD37254;  
XX  
DT 21-AUG-2002 (first entry)  
XX  
DE Adeno-associated virus (AAV) vector plasmid #1.  
XX  
KW Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;  
KW adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;  
KW Becker muscular dystrophy; ds.  
XX  
OS Chimeric - Homo sapiens.  
OS Chimeric - Unidentified.  
XX  
PN WO200183695-A2.  
XX  
PD 08-NOV-2001.  
XX  
PF 27-APR-2001; 2001WO-US13677.  
XX  
PR 28-APR-2000; 2000US-200777P.  
XX  
PA (XIAO/) XIAO X.  
XX

PI xiao X;  
XX  
DR WPI; 2002-049342/06.  
XX  
PT New dystrophin minigene for treating Duchenne or Becker muscular  
PT dystrophy comprises an N-terminal domain or modified N-terminal domain,  
PT rod repeats, H1 and H4 domains and a cysteine rich domain of a  
PT dystrophin gene -  
XX  
PS Example 1; Page 57; 71pp; English.  
XX  
CC The present invention relates to an isolated nucleotide sequence encoding  
CC a dystrophin minigene. The minigene comprises N-terminal or modified  
CC N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4  
CC domains and cysteine-rich domains of dystrophin or utrophin genes. The  
CC invention also relates to a recombinant adeno-associated virus (AAV)  
CC comprising dystrophin minigene operably linked to an expression control  
CC element. The dystrophin minigene in operable linkage with an expression  
CC control element, in a recombinant adeno-associated virus or retrovirus is  
CC useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular  
CC dystrophy (BMD) in a mammalian subject. The present sequence is AAV  
CC vector plasmid containing human dystrophin minigenes, a muscle creatine  
CC kinase (MCK) promoter and a small polyA signal sequence.  
XX  
SQ Sequence 955 BP; 177 A; 307 C; 296 G; 175 T; 0 other;

Query Match 75.1%; Score 107.4; DB 24; Length 955;  
Best Local Similarity 85.1%; Pred. No. 2.4e-19;  
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 1 TTGCCACTCCCTCTCTGCGGCTCGCTCGCTGGGGGCTGCGACCAAGTCCGC 60  
DB 1 TTGGCCACTCCCTCTCTGCGGCTCGCTCGCTGAGGCCGGGCGACCAAGTCCGC 60  
QY 61 AGACGGCAGAGCTCTGCTCTGCGGCGCCACCGAGCGAGCGCGCAGAGGAGTG 120  
DB 61 CGACGCGCGGGCTTGTGCGCCCGGCTCAGTGAGCGAGCGCGCAGAGGAGTG 120  
QY 121 GGCACTCCATCACTAGGGGT 141  
DB 121 GCCAACTCCATCACTAGGGGT 141

Search completed: July 20, 2003, 09:39:47  
Job time : 32.1207 secs





DB 61 CGACGCCCGGCTTTGGCCCGCGGCTCAGTGAGCGAGCGA3CGCGCAGAGAGGAGTG 120  
QY 121 GGCAACTCCATCCTAGGGGT 141  
DB 121 GCCAACTCCATCCTAGGGGT 141

## RESULT 2

US-08-702-573-4  
; Sequence 4, Application US/08702573  
; Patent No. 6033885  
; GENERAL INFORMATION:  
; APPLICANT: LATTI, Martine  
; APPLICANT: DENEFFLE, Patrice  
; APPLICANT: VIGNE, Emmanuelle  
; APPLICANT: PERRICAUDET, Michel  
; TITLE OF INVENTION: INTEGRATIVE RECOMBINANT ADENOVIRUSES,  
; TITLE OF INVENTION: PREPARATION THEREOF AND THERAPEUTICAL USES THEREOF  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Rhone-Poulenc Rorer Inc.  
; STREET: 500 Arcola Rd. 3C43  
; CITY: Collegeville  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19426  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/702,573  
; FILING DATE:  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: FR 94/02445  
; FILING DATE: 03-MAR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: WO PCT/FR95/00233  
; FILING DATE: 28-FEB-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Smith Ph.D., Julie K.  
; REGISTRATION NUMBER: 38,619  
; REFERENCE/DOCKET NUMBER: ST94011-US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (610)454-3839  
; TELEFAX: (610)454-3808  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 145 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: other nucleic acid  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 1..145  
; OTHER INFORMATION: /note= "Minimal ITR Sequence"  
; US-08-702-573-4

Query Match 75.1%; Score 107.4; DB 3; Length 145;  
Best Local Similarity 85.1%; Pred. No. 2.1e-21;  
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 1 TTGCCCACTCCCTCTGCGCGCTCGTGGGGCTGCGGACCAAGTCCGC 60  
DB 1 TTGGCCACTCCCTCTGCGCGCTCGTGGGGCTGCGGACCAAGTCCGC 60  
QY 61 AGACGGCAGAGCTGTCTGCGCGCGCGCGCGCGCGCGCGCGAGAGGAGTG 120  
DB 61 CGACGCCCGGCTTTGGCCCGCGGCTCAGTGAGCGAGCGCGCAGAGAGGAGTG 120

QY 121 GGCAACTCCATCCTAGGGGT 141  
DB 121 GCCAACTCCATCCTAGGGGT 141

## RESULT 3

US-07-982-193-1  
; Sequence 1, Application US/07982193  
; Patent No. 6261834  
; GENERAL INFORMATION:  
; APPLICANT: Srivastava, Arun  
; TITLE OF INVENTION: SAFE VECTOR FOR GENE THERAPY  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Scully, Scott, Murphy & Presser  
; STREET: 400 Garden City Plaza  
; CITY: Garden City  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 11530  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/982,193  
; FILING DATE: 19921125  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: McNulty, William E.  
; REGISTRATION NUMBER: 22,606  
; REFERENCE/DOCKET NUMBER: 8361  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (516) 742-4343  
; TELEFAX: (516) 742-4366  
; TELEX: 230 901 SANS UR  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 145 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; US-07-982-193-1

Query Match 75.1%; Score 107.4; DB 4; Length 145;  
Best Local Similarity 85.1%; Pred. No. 2.1e-21;  
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 1 TTGCCCACTCCCTCTGCGCGCTCGTGGGGCTGCGGACCAAGTCCGC 60  
DB 1 TTGGCCACTCCCTCTGCGCGCTCGTGGGGCTGCGGACCAAGTCCGC 60  
QY 61 AGACGGCAGAGCTGTCTGCGCGCGCGCGCGCGCGCGAGAGGAGTG 120  
DB 61 CGACGCCCGGCTTTGGCCCGCGGCTCAGTGAGCGAGCGCGCAGAGAGGAGTG 120  
QY 121 GGCAACTCCATCCTAGGGGT 141  
DB 121 GCCAACTCCATCCTAGGGGT 141

## RESULT 4

US-07-989-841A-1  
; Sequence 1, Application US/07989841A  
; Patent No. 5478745  
; GENERAL INFORMATION:  
; APPLICANT: Samulski, R. J.  
; TITLE OF INVENTION: Recombinant Viral Vector System  
; NUMBER OF SEQUENCES: 6

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/989,841A
FILING DATE: On even date herewith
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 6636-013
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 165 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
US-07-989-841A-1

```

Query Match	75.1%;	Score 107.4;	DB 1;	Length 165;
Best Local Similarity	85.1%;	Pred. No. 2.2e-21;		
Matches 120; Conservative	0;	Mismatches 21;	Indels 0;	Gaps 0;

QY	1	TTGCCACTCCCTCTCTGCGCGCTCGCTCGGTGGGGCCCTGCGGACCAAGTCCGC	60
Db	21	TTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGCGACCAAGTCCGC	80
QY	61	AGACGGCAGAGCTCTGCTCTGCGCGGCCCAACGAGCGAGCGCGCAGAGAGGAGTG	120
Db	81	CGACGCCCGGGCTTTGCCCCGGCGGCTCACTGAGCGAGCGCGCAGAGAGGAGTG	140
QY	121	GGCAACTCCATCACTAGGGGT	141
Db	141	GCCCACTCCATCACTAGGGGT	161

```

RESULT 5
US-08-440-738A-1
; Sequence 1, Application US/08440738A
; Patent No. 5869305
; GENERAL INFORMATION:
; APPLICANT: Samulski, R. J.
; APPLICANT: Xiao, X.
; TITLE OF INVENTION: Recombinant Viral Vector System
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440, 738A

```

```

; FILING DATE: May 15, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 6636-022
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 165 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
;
US-08-440-738A-1

```

Query Match	75.1%;	Score 107.4;	DB 2;	Length 165;
Best Local Similarity	85.1%;	Pred. No. 2.2e-21;		
Matches 120; Conservative	0;	Mismatches 21;	Indels 0;	Gaps 0;

[illegible]

```

RESULT 6
US-08-471-914-1
; Sequence 1, Application US/08471914A
; Patent No. 6057152
; GENERAL INFORMATION:
; APPLICANT: Samulek1, R.
; APPLICANT: xiao, X.
; TITLE OF INVENTION: RECOMBINANT VIRAL VECTOR SYSTEM
; FILE REFERENCE: 6636-027
; CURRENT APPLICATION NUMBER: US/08/471,914A
; CURRENT FILING DATE: 1995-06-06
; EARLIER APPLICATION NUMBER: 08/440,738
; EARLIER FILING DATE: 1995-05-15
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 165.
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: double-D
; US-08-471-914-1

```

[illegible]

Db 141 GCCAACTCCATCACTAGGGGT 161

## RESULT 7

US-09-276-625-7  
; Sequence 7, Application US/09276625  
; Patent No. 6436392  
; GENERAL INFORMATION:  
; APPLICANT: Engelhardt, John F.  
; TITLE OF INVENTION: Adeno-associated virus vectors  
; FILE REFERENCE: 875.007US1  
; CURRENT FILING DATE: 1999-03-25  
; PRIOR APPLICATION NUMBER: US 60/086,166  
; PRIOR FILING DATE: 1998-05-20  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 7  
; LENGTH: 165  
; TYPE: DNA  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: SEQ ID NO:1 of U.S. Patent No. 6436392 5,478,745  
US-09-276-625-7

Query Match 75.1%; Score 107.4; DB 4; Length 165;  
Best Local Similarity 85.1%; Pred. No. 2.2e-21;  
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 1 TTGCCACTCCCTCTCTGCGCGCTCGCTCGCTCGGTGGGGCTGCGACCAAGTCCGC 60  
|||  
Db 21 TTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGCGCGGCGACCAAGTCCGC 80  
|||  
QY 61 AGACGGCAGAGCTGTGCTCTGCGCGCGCGCGCGCGAGCGAGCGCGAGAGGAGTG 120  
|||  
Db 81 CGACGCGCGGCTTGGCGCGCGCGCGCGCTCACTGAGCGAGCGCGCGAGAGGAGTG 140  
|||  
QY 121 GGCACTCCATCACTAGGGGT 141  
|||  
Db 141 GCCAACTCCATCACTAGGGGT 161  
|||

## RESULT 8

US-09-276-625-6  
; Sequence 6, Application US/09276625  
; Patent No. 6436392  
; GENERAL INFORMATION:  
; APPLICANT: Engelhardt, John F.  
; TITLE OF INVENTION: Adeno-associated virus vectors  
; FILE REFERENCE: 875.007US1  
; CURRENT FILING DATE: 1999-03-25  
; PRIOR APPLICATION NUMBER: US 60/086,166  
; PRIOR FILING DATE: 1998-05-20  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6  
; LENGTH: 272  
; TYPE: DNA  
; ORGANISM: AAV circular intermediate, clone p1202  
US-09-276-625-6

Query Match 75.1%; Score 107.4; DB 4; Length 272;  
Best Local Similarity 85.1%; Pred. No. 2.3e-21;  
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 1 TTGCCACTCCCTCTCTGCGCGCTCGCTCGCTCGGTGGGGCTGCGACCAAGTCCGC 60  
|||  
Db 69 TTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGCGCGGCGAGCAAAAGTCCGC 128  
|||

QY 61 AGACGGCAGAGCTGTGCTCTGCGCGCGCGCGCGAGCGAGCGCGCGAGAGGAGTG 120  
|||  
Db 129 CGACGCGCGGCTTGGTTCGCCCGGCTCACTGAGCGAGCGCGCGAGAGGAGTG 188  
|||  
QY 121 GGCACTCCATCACTAGGGGT 141  
|||  
Db 189 GCCAACTCCATCACTAGGGGT 209  
|||

## RESULT 9

US-08-254-358-1  
; Sequence 1, Application US/08254358  
; Patent No. 5658785  
; GENERAL INFORMATION:  
; APPLICANT: Johnson, Phillip R.  
; TITLE OF INVENTION: Adeno-Associated Virus Materials and  
; TITLE OF INVENTION: Methods  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 S. Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/254,358  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 5658785and, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 31975  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 474-6300  
TELEFAX: (312) 474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4680 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-254-358-1

Query Match 75.1%; Score 107.4; DB 1; Length 4680;  
Best Local Similarity 85.1%; Pred. No. 3e-21;  
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 1 TTGCCACTCCCTCTCTGCGCGCTCGCTCGCTCGGTGGGGCTGCGACCAAGTCCGC 60  
|||  
Db 1 TTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGCGCGGCGACCAAGTCCGC 60  
|||  
QY 61 AGACGGCAGAGCTGTGCTCTGCGCGCGCGCGCGAGCGAGCGCGCGAGAGGAGTG 120  
|||  
Db 61 CGACGCGCGGCTTGGCGCGCGCGCGCGCTCACTGAGCGAGCGCGCGAGAGGAGTG 120  
|||  
QY 121 GGCACTCCATCACTAGGGGT 141  
|||  
Db 121 GCCAACTCCATCACTAGGGGT 141  
|||

## RESULT 10

US-08-475-391-1  
; Sequence 1, Application US/08475391  
; Patent No. 5786211  
; GENERAL INFORMATION:



APPLICANT: Johnson, Philip R.  
TITLE OF INVENTION: Adeno-Associated Virus Materials and  
TITLE OF INVENTION: Methods  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 S. Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/475,391  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/254,358  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 578621and, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 31975  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 474-6300  
TELEFAX: (312) 474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4680 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-475-391-1

Query Match 75.1%; Score 107.4; DB 1; Length 4680;  
Best Local Similarity 85.1%; Pred. No. 3e-21;  
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 1 TTGCCCACTCCCTCTCTGCGCGCTCGCTCGCTGCGGCGCTCGGACCAAGTCCGC 60  
|||  
DB 1 TTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGCGCGCGGACCAAGTCCGC 60  
QY 61 AGACGGCAGAGCTCTGCTGCGCGCGCGCGCGCGAGCGAGCGCGCGCAGAGAGGAGTG 120  
|||  
DB 61 CGACGCCCGGGCTTTGCCCGCGCGCGCTCTCAGTGAGCGAGCGAGCGCGCAGAGAGGAGTG 120  
QY 121 GGCACTCCATCACTAGGGGT 141  
|||  
DB 121 GCCAACTCCATCACTAGGGGT 141

RESULT 11  
US-08-709-609-1  
Sequence 1, Application US/08709609  
Patent No. 5858775  
GENERAL INFORMATION:  
APPLICANT: Johnson, Philip R.  
TITLE OF INVENTION: Adeno-Associated Virus Materials and  
TITLE OF INVENTION: Methods  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 S. Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/709,609  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 5858775and, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 31975  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 474-6300  
TELEFAX: (312) 474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4680 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-709-609-1

Query Match 75.1%; Score 107.4; DB 2; Length 4680;  
Best Local Similarity 85.1%; Pred. No. 3e-21;  
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 1 TTGCCCACTCCCTCTCTGCGCGCTCGCTCGCTGCGGCGCTCGGACCAAGTCCGC 60  
|||  
DB 1 TTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGCGCGCGGACCAAGTCCGC 60  
QY 61 AGACGGCAGAGCTCTGCTGCGCGCGCGCGCGCGAGCGAGCGCGCGCAGAGAGGAGTG 120  
|||  
DB 61 CGACGCCCGGGCTTTGCCCGCGCGCGCTCTCAGTGAGCGAGCGAGCGCGCAGAGAGGAGTG 120  
QY 121 GGCACTCCATCACTAGGGGT 141  
|||  
DB 121 GCCAACTCCATCACTAGGGGT 141

RESULT 12  
PCT-US95-07178-1  
Sequence 1, Application PC/TUS9507178  
GENERAL INFORMATION:  
APPLICANT: Johnson, Philip R.  
TITLE OF INVENTION: Adeno-Associated Virus Materials and  
TITLE OF INVENTION: Methods  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 S. Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/07178  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Noland, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 31975  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 474-6300

TELEFAX: (312) 474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4680 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
PCT-US95-07178-1

Query Match 75.1%; Score 107.4; DB 5; Length 4680;  
Best Local Similarity 85.1%; Pred. No. 3e-21;  
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 1 TTGCCACTCCCTCTCTGCGCGCTCGCTCGGTGGGCGCTGCGGACCAAGTCCGC 60  
DB 1 TTGGCCACTCCCTCTCTGCGCGCTCGCTCGTCACTGAGCGCGGCGACCAAGTCCGC 60  
QY 61 AGACGGCAGAGCTCTGCTCTGCGCGCGCCACCGAGCGAGCGCGCAGAGGAGTGTG 120  
DB 61 CGACGCGCGCGCTTGGCGCGCGCGCTCAGTAGAGCGAGCGCGCAGAGGAGTGTG 120  
QY 121 GGCACTCCATCACTAGGGGT 141  
DB 121 GCCAACTCCATCACTAGGGGT 141

RESULT 13  
US-09-299-141-4  
Sequence 4, Application US/09299141  
Patent No. 6461606  
GENERAL INFORMATION:  
APPLICANT: FLOTTE, TERENCE R.  
APPLICANT: SONG, SIHONG  
APPLICANT: BYRNE, BARRY J.  
APPLICANT: MORGAN, MICHAEL  
TITLE OF INVENTION: MATERIALS AND METHODS FOR GENE THERAPY  
FILE REFERENCE: 4300.011800  
CURRENT APPLICATION NUMBER: US/09/299,141  
CURRENT FILING DATE: 1999-04-23  
EARLIER APPLICATION NUMBER: 60/083,025  
EARLIER FILING DATE: 1998-04-24  
NUMBER OF SEQ ID NOS: 13  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 4  
LENGTH: 5932  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence:p43C-AT  
US-09-299-141-4

Query Match 75.1%; Score 107.4; DB 4; Length 5932;  
Best Local Similarity 85.1%; Pred. No. 3.1e-21;  
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 1 TTGCCACTCCCTCTCTGCGCGCTCGCTCGGTGGGCGCTGCGGACCAAGTCCGC 60  
DB 18 TTGGCCACTCCCTCTCTGCGCGCTCGCTCGTCACTGAGCGCGGCGACCAAGTCCGC 77  
QY 61 AGACGGCAGAGCTCTGCTCTGCGCGCGCCACCGAGCGAGCGCGCAGAGGAGTGTG 120  
DB 78 CGACGCGCGCGCTTGGCGCGCGCGCTCAGTAGAGCGAGCGCGCAGAGGAGTGTG 137  
QY 121 GGCACTCCATCACTAGGGGT 141  
DB 138 GCCAACTCCATCACTAGGGGT 158

RESULT 14  
US-09-299-141-4/c  
Sequence 4, Application US/09299141

Patent No. 6461606  
GENERAL INFORMATION:  
APPLICANT: FLOTTE, TERENCE R.  
APPLICANT: SONG, SIHONG  
APPLICANT: BYRNE, BARRY J.  
APPLICANT: MORGAN, MICHAEL  
TITLE OF INVENTION: MATERIALS AND METHODS FOR GENE THERAPY  
FILE REFERENCE: 4300.011800  
CURRENT APPLICATION NUMBER: US/09/299,141  
CURRENT FILING DATE: 1999-04-23  
EARLIER APPLICATION NUMBER: 60/083,025  
EARLIER FILING DATE: 1998-04-24  
NUMBER OF SEQ ID NOS: 13  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 4  
LENGTH: 5932  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence:p43C-AT  
US-09-299-141-4

Query Match 75.1%; Score 107.4; DB 4; Length 5932;  
Best Local Similarity 85.1%; Pred. No. 3.1e-21;  
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 1 TTGCCACTCCCTCTCTGCGCGCTCGCTCGGTGGGCGCTGCGGACCAAGTCCGC 60  
DB 3078 TTGGCCACTCCCTCTCTGCGCGCTCGCTCGTCACTGAGCGCGGCGACCAAGTCCGC 3019  
QY 61 AGACGGCAGAGCTCTGCTCTGCGCGCGCCACCGAGCGAGCGCGCAGAGGAGTGTG 120  
DB 3018 CGACGCGCGCGCTTGGCGCGCGCGCTCAGTAGAGCGAGCGCGCAGAGGAGTGTG 2959  
QY 121 GGCACTCCATCACTAGGGGT 141  
DB 2958 GCCAACTCCATCACTAGGGGT 2938

RESULT 15  
US-09-299-141-8  
Sequence 8, Application US/09299141  
Patent No. 6461606  
GENERAL INFORMATION:  
APPLICANT: FLOTTE, TERENCE R.  
APPLICANT: SONG, SIHONG  
APPLICANT: BYRNE, BARRY J.  
APPLICANT: MORGAN, MICHAEL  
TITLE OF INVENTION: MATERIALS AND METHODS FOR GENE THERAPY  
FILE REFERENCE: 4300.011800  
CURRENT APPLICATION NUMBER: US/09/299,141  
CURRENT FILING DATE: 1999-04-23  
EARLIER APPLICATION NUMBER: 60/083,025  
EARLIER FILING DATE: 1998-04-24  
NUMBER OF SEQ ID NOS: 13  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 8  
LENGTH: 6142  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence:PLASMID  
US-09-299-141-8

Query Match 75.1%; Score 107.4; DB 4; Length 6142;  
Best Local Similarity 85.1%; Pred. No. 3.1e-21;  
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 1 TTGCCACTCCCTCTCTGCGCGCTCGCTCGGTGGGCGCTGCGGACCAAGTCCGC 60  
DB 18 TTGGCCACTCCCTCTCTGCGCGCTCGCTCGTCACTGAGCGCGGCGACCAAGTCCGC 77

Qy 61 AGACGGCAGAGCTCTGCTCTGCCCCCACCAGAGCGAGCGCGCCAGAGAGGGAGTG 120  
Db 78 CGACGCCCGGGCTTTGCCCGGGCGGCTCAGTAGAGCGAGCGCGCCAGAGAGGGAGTG 137  
Qy 121 GGCACTCCATCACTAGGGT 141  
Db 138 GCCAACTCCATCACTAGGGT 158

Search completed: July 20, 2003, 13:23:48  
Job time : 6.68705 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2003, 13:19:56 ; Search time 30.7204 Seconds  
(without alignments)  
9603.029 Million cell updates/sec

Title: US-09-807-802A-1\_COPY\_1\_143  
Perfect score: 143  
Sequence: 1 ttgccacatccctctctgcg.....aactccatcactagggttaa 143

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1439767 seqs, 1031500376 residues

Total number of hits satisfying chosen parameters: 2879534

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA:\*

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq1:\*  
11: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq2:\*  
12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq3:\*  
13: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*  
15: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
16: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
17: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	107.4	75.1	145	11	US-09-782-378A-6
2	107.4	75.1	146	15	US-10-135-98A-8
3	107.4	75.1	165	11	US-09-782-378A-8
4	107.4	75.1	165	15	US-10-054-665-7
5	107.4	75.1	207	15	US-10-023-208-58
6	107.4	75.1	272	15	US-10-054-665-6
7	107.4	75.1	4675	11	US-09-782-378A-1
8	107.4	75.1	4675	11	US-09-782-378A-2
9	107.4	75.1	4675	15	US-10-240-198-1
10	107.4	75.1	4679	10	US-09-804-898-1
11	107.4	75.1	4679	10	US-09-945-681-10
12	107.4	75.1	4679	15	US-10-038-972A-12
13	107.4	75.1	4680	15	US-10-077-294-1
14	107.4	75.1	4680	15	US-10-163-886-1
15	107.4	75.1	4680	15	US-10-263-127-1
16	107.4	75.1	5932	15	US-10-267-117-4

C	17	107.4	75.1	5932	15	US-10-267-117-4	Sequence 4, Appli
C	18	107.4	75.1	5932	15	US-10-340-112-4	Sequence 4, Appli
C	19	107.4	75.1	5932	15	US-10-340-112-4	Sequence 4, Appli
C	20	107.4	75.1	6142	15	US-10-267-117-8	Sequence 8, Appli
C	21	107.4	75.1	6142	15	US-10-267-117-8	Sequence 8, Appli
C	22	107.4	75.1	6142	15	US-10-340-112-8	Sequence 8, Appli
C	23	107.4	75.1	6142	15	US-10-340-112-8	Sequence 8, Appli
C	24	107.4	75.1	6565	15	US-10-267-117-1	Sequence 1, Appli
C	25	107.4	75.1	6565	15	US-10-267-117-1	Sequence 1, Appli
C	26	107.4	75.1	6565	15	US-10-340-112-1	Sequence 1, Appli
C	27	107.4	75.1	6565	15	US-10-340-112-1	Sequence 1, Appli
C	28	107.4	75.1	6714	15	US-10-267-117-6	Sequence 6, Appli
C	29	107.4	75.1	6714	15	US-10-267-117-6	Sequence 6, Appli
C	30	107.4	75.1	6714	15	US-10-340-112-6	Sequence 6, Appli
C	31	107.4	75.1	6714	15	US-10-340-112-6	Sequence 6, Appli
C	32	107.4	75.1	6924	15	US-10-267-117-9	Sequence 9, Appli
C	33	107.4	75.1	6924	15	US-10-267-117-9	Sequence 9, Appli
C	34	107.4	75.1	6924	15	US-10-267-117-10	Sequence 10, Appli
C	35	107.4	75.1	6924	15	US-10-267-117-10	Sequence 10, Appli
C	36	107.4	75.1	6924	15	US-10-267-117-11	Sequence 11, Appli
C	37	107.4	75.1	6924	15	US-10-267-117-11	Sequence 11, Appli
C	38	107.4	75.1	6924	15	US-10-340-112-9	Sequence 9, Appli
C	39	107.4	75.1	6924	15	US-10-340-112-9	Sequence 9, Appli
C	40	107.4	75.1	6924	15	US-10-340-112-10	Sequence 10, Appli
C	41	107.4	75.1	6924	15	US-10-340-112-10	Sequence 10, Appli
C	42	107.4	75.1	6924	15	US-10-340-112-11	Sequence 11, Appli
C	43	107.4	75.1	6924	15	US-10-340-112-11	Sequence 11, Appli
C	44	107.4	75.1	6981	15	US-10-267-117-7	Sequence 7, Appli
C	45	107.4	75.1	6981	15	US-10-267-117-7	Sequence 7, Appli

## ALIGNMENTS

RESULT 1  
US-09-782-378A-6  
; Sequence 6, Application US/09782378A  
; Patent No. US20020102731A1  
; GENERAL INFORMATION:  
; APPLICANT: Hearing, Patrick  
; APPLICANT: Bahou, Wadie  
; APPLICANT: Sandalon, Ziv  
; APPLICANT: Gnatenko, Dmitri  
; TITLE OF INVENTION: Adenoviral Vectors  
; FILE REFERENCE: STONYB-04970  
; CURRENT APPLICATION NUMBER: US/09/782,378A  
; PRIOR APPLICATION NUMBER: 2001-02-12  
; PRIOR FILING DATE: 2000-10-02  
; NUMBER OF SEQ ID NOS: 27  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 6  
; LENGTH: 145  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-782-378A-6

Query Match 75.1%; Score 107.4; DB 11; Length 145;  
Best Local Similarity 85.1%; Pred. No. 2.8e-21;  
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;  
QY 1 TTGCCACTCCCTCTGCGCGCTCGCTCGGTGGGGCTGCGACCAAGTCCGC 60  
DB 1 TTGCCACTCCCTCTGCGCGCTCGCTCGTCACTAGAGCGCGGCGACCAAGTCCGC 60  
QY 61 AGACGGCAGAGCTCTGCTCTGCGCGCCACCGAGCGAGCGAGCGAGGAGTG 120  
DB 61 CGACGCGCGGCTTTGCCCGCGCGCTCACTAGAGCGAGCGAGCGAGGAGTG 120  
QY 121 GGCACTTCATCACTAGGGGT 141  
DB 121 GCCAATTCATCACTAGGGGT 141







;; CURRENT APPLICATION NUMBER: US/10/240,198  
;; CURRENT FILING DATE: 2002-09-30  
;; PRIOR APPLICATION NUMBER: 0009887.1  
;; PRIOR FILING DATE: 2000-04-20  
;; NUMBER OF SEQ ID NOS: 6  
;; SOFTWARE: Patentin Ver. 2.1  
;; SEQ ID NO 1  
;; LENGTH: 4675  
;; TYPE: DNA  
;; ORGANISM: adeno-associated virus 2  
US-10-240-198-1

Query Match 75.1%; Score 107.4; DB 15; Length 4675;  
Best Local Similarity 85.1%; Pred. No. 1.5e-21;  
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 1 TTGCCCACTCCCTCTCTGCGCGCTCGCTCGGTGGGGCTTGGACCAAGTCCGC 60  
DB 1 TTGGCCACTCCCTCTCTGCGCGCTCGCTCGGTGAGCGCGGCGCAAAAGTCCGC 60  
QY 61 AGACGGCAGAGCTGTGCTGTGCGCGCGCGCGCGAGCGAGCGCGCGAGAGGAGTG 120  
DB 61 CGACGCGCGGCTTTGCGCGCGCGCGCTCAGTGAGCGAGCGCGCGAGAGGAGTG 120  
QY 121 GGCACTCCATCACTAGGGGT 141  
DB 121 GCCAACTCCATCACTAGGGGT 141

RESULT 10  
US-09-804-898-1  
; Sequence 1, Application US/09804898  
; Patent No. US20020045264A1  
; GENERAL INFORMATION:  
; APPLICANT: DURING, MATTHEW  
; APPLICANT: XIAO, WEIDONG  
; TITLE OF INVENTION: PRODUCTION OF CHIMERIC CAPSID VECTORS  
; FILE REFERENCE: 102182-14  
; CURRENT APPLICATION NUMBER: US/09/804,898  
; CURRENT FILING DATE: 2001-03-13  
; PRIOR APPLICATION NUMBER: 60/1189,110  
; PRIOR FILING DATE: 2000-03-14  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: Patentin version 3.0  
; SEQ ID NO 1  
; LENGTH: 4679  
; TYPE: DNA  
; ORGANISM: adeno-associated virus 2  
US-09-804-898-1

Query Match 75.1%; Score 107.4; DB 10; Length 4679;  
Best Local Similarity 85.1%; Pred. No. 1.5e-21;  
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 1 TTGCCCACTCCCTCTCTGCGCGCTCGCTCGGTGGGGCTTGGACCAAGTCCGC 60  
DB 1 TTGGCCACTCCCTCTCTGCGCGCTCGCTCGGTGAGCGCGGCGCAAAAGTCCGC 60  
QY 61 AGACGGCAGAGCTGTGCTGTGCGCGCGCGCGAGCGAGCGCGCGAGAGGAGTG 120  
DB 61 CGACGCGCGGCTTTGCGCGCGCGCGCTCAGTGAGCGAGCGCGCGAGAGGAGTG 120  
QY 121 GGCACTCCATCACTAGGGGT 141  
DB 121 GCCAACTCCATCACTAGGGGT 141

RESULT 11  
US-09-945-681-10  
; Sequence 10, Application US/09945681  
; Patent No. US20020064878A1  
; GENERAL INFORMATION:  
; APPLICANT: UNIVERSITE DE NANTES

;; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR RECOMBINANT  
;; TITLE OF INVENTION: ADENO-ASSOCIATED VIRUS PRODUCTION  
;; FILE REFERENCE: B4182AA - UNIVERSITE DE NANTES  
;; CURRENT APPLICATION NUMBER: US/09/945,681  
;; CURRENT FILING DATE: 2001-09-05  
;; PRIOR APPLICATION NUMBER: PCT/EP 00/01854  
;; PRIOR FILING DATE: 2000-03-05  
;; NUMBER OF SEQ ID NOS: 10  
;; SOFTWARE: Patentin Ver. 2.1  
;; SEQ ID NO 10  
;; LENGTH: 4679  
;; TYPE: DNA  
;; ORGANISM: adeno-associated virus 2  
US-09-945-681-10

Query Match 75.1%; Score 107.4; DB 10; Length 4679;  
Best Local Similarity 85.1%; Pred. No. 1.5e-21;  
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 1 TTGCCCACTCCCTCTCTGCGCGCTCGCTCGGTGGGGCTTGGACCAAGTCCGC 60  
DB 1 TTGGCCACTCCCTCTCTGCGCGCTCGCTCGGTGAGCGCGGCGCAAAAGTCCGC 60  
QY 61 AGACGGCAGAGCTGTGCTGTGCGCGCGCGCGAGCGAGCGCGCGAGAGGAGTG 120  
DB 61 CGACGCGCGGCTTTGCGCGCGCGCGCTCAGTGAGCGAGCGCGCGAGAGGAGTG 120  
QY 121 GGCACTCCATCACTAGGGGT 141  
DB 121 GCCAACTCCATCACTAGGGGT 141

RESULT 12  
US-10-038-972A-12  
; Sequence 12, Application US/10038972A  
; Publication No. US20020192823A1  
; GENERAL INFORMATION:  
; APPLICANT: J. Bartlett  
; TITLE OF INVENTION: AAV VECTORS AND METHODS  
; FILE REFERENCE: 28335/36996US  
; CURRENT APPLICATION NUMBER: US/10/038,972A  
; CURRENT FILING DATE: 2002-01-04  
; PRIOR APPLICATION NUMBER: US 60/260,124  
; PRIOR FILING DATE: 2001-01-05  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 12  
; LENGTH: 4679  
; TYPE: DNA  
; ORGANISM: adeno-associated virus 2  
US-10-038-972A-12

Query Match 75.1%; Score 107.4; DB 15; Length 4679;  
Best Local Similarity 85.1%; Pred. No. 1.5e-21;  
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 1 TTGCCCACTCCCTCTCTGCGCGCTCGCTCGGTGGGGCTTGGACCAAGTCCGC 60  
DB 1 TTGGCCACTCCCTCTCTGCGCGCTCGCTCGGTGAGCGCGGCGCAAAAGTCCGC 60  
QY 61 AGACGGCAGAGCTGTGCTGTGCGCGCGCGCGAGCGAGCGCGCGAGAGGAGTG 120  
DB 61 CGACGCGCGGCTTTGCGCGCGCGCGCTCAGTGAGCGAGCGCGCGAGAGGAGTG 120  
QY 121 GGCACTCCATCACTAGGGGT 141  
DB 121 GCCAACTCCATCACTAGGGGT 141

RESULT 13  
US-10-077-294-1  
; Sequence 1, Application US/10077294  
; Publication No. US20020159979A1

```
;; GENERAL INFORMATION:
;; APPLICANT: Johnson, Philip R.
;; TITLE OF INVENTION: Adeno-Associated Virus Materials and
;; Methods
;; NUMBER OF SEQUENCES: 3
;; CORRESPONDENCE ADDRESS:
;; ADDRESS: Marshall, O'Toole, Gerstein, Murray & Borun
;; STREET: 6300 Sears Tower, 233 S. Wacker Drive
;; CITY: Chicago
;; STATE: Illinois
;; COUNTRY: USA
;; ZIP: 60606
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentln Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/10/077,294
;; FILING DATE: 15-Feb-2002
;; CLASSIFICATION: <Unknown>
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 09/691,604
;; FILING DATE: <Unknown>
;; ATTORNEY/AGENT INFORMATION:
;; NAME: No. US20020159979A1and, Greta E.
;; REGISTRATION NUMBER: 35,302
;; REFERENCE/DOCKET NUMBER: 31975
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (312) 474-6300
;; TELEFAX: (312) 474-0448
;; TELEX: 25-3856
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 4680 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
;; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-10-077-294-1

Query Match      75.1%; Score 107.4; DB 15; Length 4680;
Best Local Similarity 85.1%; Pred.No.1.5e-21;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 1 TTGCCACTCCCTCTCTGCGCGCTCGCTCGCTCGGTGGGGCTGCGACCAAGTCCGC 60
   |||
Db 1 TTGGCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGCGCGGCGCAACCAAGTCCGC 60

QY 61 AGACGGCAGAGCTCTGCTGTGCGCGCCGCCACCGAGCGAGCGCGCGCAGAGAGGAGTG 120
   |||
Db 61 CGACGCCCGGGCTTTGCCCCGGGCGCTCAGTGAGCGAGCGCGCGCAGAGAGGAGTG 120

QY 121 GGCACTCCATCACTAGGGGT 141
   |||
Db 121 GCCAATCCATCACTAGGGGT 141

RESULT 14
US-10-163-886-1
; Sequence 1, Application US/10163886
; Publication No. US20020187129A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Philip R.
; TITLE OF INVENTION: Adeno-Associated Virus Materials and
; Methods
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESS: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 S. Wacker Drive
; CITY: Chicago
; STATE: Illinois
```

```
;; COUNTRY: USA
;; ZIP: 60606
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentln Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/10/163,886
;; FILING DATE: 04-Jun-2002
;; CLASSIFICATION: <Unknown>
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 09/292,703
;; FILING DATE: <Unknown>
;; ATTORNEY/AGENT INFORMATION:
;; NAME: No. US20020187129A1and, Greta E.
;; REGISTRATION NUMBER: 35,302
;; REFERENCE/DOCKET NUMBER: 31975
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (312) 474-6300
;; TELEFAX: (312) 474-0448
;; TELEX: 25-3856
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 4680 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
;; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-10-163-886-1

Query Match      75.1%; Score 107.4; DB 15; Length 4680;
Best Local Similarity 85.1%; Pred.No.1.5e-21;
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
```

```
QY 1 TTGCCACTCCCTCTCTGCGCGCTCGCTCGCTCGGTGGGGCTGCGACCAAGTCCGC 60
   |||
Db 1 TTGGCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGCGCGGCGCAACCAAGTCCGC 60

QY 61 AGACGGCAGAGCTCTGCTGTGCGCGCCGCCACCGAGCGAGCGCGCGCAGAGAGGAGTG 120
   |||
Db 61 CGACGCCCGGGCTTTGCCCCGGGCGCTCAGTGAGCGAGCGCGCGCAGAGAGGAGTG 120

QY 121 GGCACTCCATCACTAGGGGT 141
   |||
Db 121 GCCAATCCATCACTAGGGGT 141

RESULT 15
US-10-263-127-1
; Sequence 1, Application US/10263127
; Publication No. US20030082145A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Philip R.
; TITLE OF INVENTION: Adeno-Associated Virus Materials and
; Methods
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESS: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 S. Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/263,127
; FILING DATE: 02-Oct-2002
```



CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/10/077,294

FILING DATE: 15-Feb-2002

APPLICATION NUMBER: 09/691,604

FILING DATE: <Unknown>

**ATTORNEY/AGENT INFORMATION:**

NAME: NO. US20030082145A1and, Greta E.

REGISTRATION NUMBER: 35,302

REFERENCE/DOCKET NUMBER: 31975

TELECOMMUNICATION INFORMATION:

TELEPHONE: (312) 474-6300

TELEFAX: (312) 474-0448

TELEX: 25-3856

INFORMATION FOR SEQ ID NO: 1:

### SEQUENCE CHARACTERISTICS:

LENGTH: 4680 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SEQUENCE	DESCRIPTION:	SEQ ID NO: 1
100000000	100000000	100000000

US-10-263-127-1

Query Match	75.1%	Score 107.4	DB 15	Length 4680
-------------	-------	-------------	-------	-------------

-Best Local Similarity 85.1%; Pred. No. 1.5e-21;

```
Matches 120; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
```

1 TGCCCACTCCCTCTCTGCGCGCTCGCTCGGTGGGGCTGCGGACCAAGTCCG 60

Db 1 TTGGCCAOTCCCTCTCTGCGCGCTCGCTCACTGAGGCCGGGGCAACCAAGTGGCC 60

61 AGACGCGAGACTCTGCTCTGCCGCCACCGAGCGAGCGAGCGCGCAGAGAGGAGTG 120

Db 61 CGACGCCCGGCTTTGCCCGGGCGGCTCAGTGAGCGAGCGGCGCAGAGAGGAGTG 120

QY 121 GGCACTCCATCACTAGGGT 141

Db 121 GCCAATCCATCACTAGGGT 141

Search completed: July 20, 2003, 20:11:28  
Job time : 31.7204 secs

Job time : 31.7204 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2003, 09:15:11 ; Search time 212.5 Seconds  
(without alignments)  
10898.630 Million cell updates/sec

Title: US-09-807-802A-1\_COPY\_1\_143  
Perfect score: 143  
Sequence: 1 ttgcccactccctctctgcg.....aactccactcactaggggttaa 143

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estcin:\*  
4: em\_estmu:\*  
5: em\_estcov:\*  
6: em\_estcpl:\*  
7: em\_estro:\*  
8: em\_hnc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hnc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	38.2	26.7	763	12	BG490707 602520211
2	37.4	26.2	644	10	BB632576 BB632576
3	36.2	25.3	920	12	BG764368 BG764368
4	36.2	25.3	932	17	CNS00720 AL066742 Drosophila
5	35.8	25.0	604	13	BI531460 BI531460
6	35.6	24.9	578	10	BB651469 BB651469

7	35.4	24.8	860	12	BF578052	BF578052 602094705
8	35.4	24.8	1123	17	AG080476	AG080476 Pan trog1
9	35.2	24.6	2024	17	AG074763	AG074763 Pan trog1
10	35	24.5	844	17	CNS0052P	AL056652 Drosophila
C 11	34.4	24.1	745	12	BG837739	BG837739 Zm10_04h1
12	34.4	24.1	890	17	AG127787	AG127787 Pan trog1
13	34	23.8	623	10	BB613613	BB613613 BB613613
14	34	23.8	639	10	BB651977	BB651977 BB651977
15	34	23.8	643	10	BB652152	BB652152 BB652152
16	34	23.8	657	10	BB651054	BB651054 BB651054
17	34	23.8	661	10	BB621081	BB621081 BB621081
C 18	34	23.8	795	17	AQ331299	AQ331299 nbxb0049M
19	33.8	23.6	881	17	CNS031ZR	AL246240 Tetraodon
C 20	33.8	23.6	932	17	CNS0072Q	AL066742 Drosophila
C 21	33.8	23.6	1136	13	BM416106	BM416106 OP21191 M
C 22	33.8	23.6	1372	13	BI913462	BI913462 603179075
23	33.6	23.5	984	12	BG809678	BG809678 mgct001xe
24	33.6	23.5	1020	17	CNS004YJ	AL055391 Drosophila
25	33.6	23.5	1136	13	BM416106	BM416106 OP21191 M
26	33.4	23.4	265	14	T14763	T14763 5c02a01-t7
27	33.4	23.4	1009	17	CNS010EW	AL098882 Drosophila
C 28	33.4	23.4	1284	13	BM547577	BM547577 AGENCOURT
C 29	33.2	23.2	827	17	AZ198560	AZ198560 SP_1037 B
30	33.2	23.2	879	17	AG141413	AG141413 Pan trog1
31	32.8	22.9	925	17	CNS0091P	AL053013 Drosophila
C 32	32.6	22.8	437	13	BM487026	BM487026 pgm2n.pk0
C 33	32.6	22.8	727	13	BI454382	BI454382 603170730
34	32.6	22.8	877	17	AG127610	AG127610 Pan trog1
C 35	32.6	22.8	1014	17	CNS0143J	AL103657 Drosophila
C 36	32.4	22.7	337	10	BE110673	BE110673 UI-R-BJ1-
C 37	32.4	22.7	445	9	AL514067	AL514067 AL514067
C 38	32.4	22.7	749	12	BG837752	BG837752 Zm10_05e0
39	32.4	22.7	822	17	AG134283	AG134283 Pan trog1
40	32.4	22.7	1100	17	AG161988	AG161988 Pan trog1
C 41	32.2	22.5	529	12	BE967846	BE967846 601643087
C 42	32.2	22.5	604	17	AQ575796	AQ575796 nbxb0087P
C 43	32.2	22.5	657	17	AG060145	AG060145 Pan trog1
44	32.2	22.5	664	17	AQ259032	AQ259032 nbxb0022E
45	32.2	22.5	692	17	AQ362958	AQ362958 nbxb0051E

ALIGNMENTS

RESULT 1  
LOCUS BG490707 763 bp mRNA EST 27-MAR-2001  
DEFINITION 60252021F1 NIH\_MGC\_18 Homo sapiens CDNA clone IMAGE:4638699 5', mRNA sequence.  
ACCESSION BG490707  
VERSION BG490707.1 GI:13452217  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 763)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: DCTD/DTP/Gazdar  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov  
Plate: LICM1401 row: e column: 04  
High quality sequence stop: 182.

FEATURES  
source 1..763







mid-log phase in TAP (acetate-containing) medium in the light, TAP medium in the dark, HS (minimal) medium in ambient levels of CO<sub>2</sub> and HS medium bubbled with 5% CO<sub>2</sub>. PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda ZAP II (Stratagene) in the EcoRI (5') and XhoI (3') sites. Bluescript II SK- plasmids were excised from the lambda ZAP clones by superinfection with ExAssist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al (1996) Genome Research 6: 791-806."

BASE COUNT  
ORIGIN

141 a 181 c 213 g 69 t

Query Match 25.0%; Score 35.8; DB 13; Length 604;  
Best Local Similarity 59.2%; Pred. No. 21;  
Matches 61; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 8 CTCCTCTCTGCGGCTCGCTCGGTGGGGCTGCGACCAAGTCCGACAGCGC 67  
DB 382 CTTCCTCTGCGCCCATCTCCGTGGATGCGCCTGCGGCTCGCCGCGCGCC 323  
QY 68 AGAGCTGTGCTGCGGCGCCACCGAGCGAGCGCGCCAG 110  
DB 322 AGCGCCCGGCTCGCGCTCTTGCGCGCTCGCGAGCTG 280

RESULT 6

BB651469

LOCUS BB651469 578 bp mRNA linear EST 26-OCT-2001  
DEFINITION BB651469 RIKEN full-length enriched, 0 day neonate cerebellum Mus  
musculus cDNA clone C230083H05 5', mRNA sequence.

ACCESSION BB651469  
VERSION BB651469.1 GI:16485722

KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 578)  
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,  
Hiramoto,K., Horii,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,  
'M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,  
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,  
D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,  
Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,  
Muramatsu,M. and Hayashizaki,Y.

TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)  
JOURNAL Unpublished (2001)  
COMMENT Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gs.c.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/

Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,  
'M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,B.,  
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,  
'S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and  
Hayashizaki,Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)  
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,  
'Y. and Hayashizaki,Y.  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a

nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa,  
'K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and  
Hayashizaki,Y.  
Computational Analysis of Full-length Mouse cDNAs Compared with  
Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)  
Please visit our web site (http://genome.gsc.riken.go.jp) for  
further details.  
e mouse tissues.

FEATURES  
Source

Location/Qualifiers  
1. 578  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="C230083H05"  
/clone\_lib="RIKEN full-length enriched, 0 day neonate  
cerebellum"  
/tissue\_type="cerebellum"  
/dev\_stage="0 day neonate"  
/lab\_host="DH10B"  
/note="Site 1: SalI; Site 2: BamHI; cDNA library was  
prepared and sequenced in Mouse Genome Encyclopedia  
Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in  
RIKEN. Division of Experimental Animal Research in Riken  
contributed to prepare mouse tissues. 1st strand cDNA was  
primed with a primer [5'  
GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTTCTTTVN 3'], cDNA was  
prepared by using trehalose thermo-activated reverse  
transcriptase and subsequently enriched for full-length by  
cap-trapper. cDNA went through one round of normalization  
to Rot = 20.0 and subtraction with the primer adapter of  
strand cDNA was prepared with the primer adapter of  
sequence [5' GAGAGAGAGATTCGAGTTAATTAATTAATCCCGCCCCCCCC  
3']. cDNA was cleaved with XhoI and BamHI. Vector: a  
modified pBluescript KS(+) after bulk excision from Lambda  
FLC I."

BASE COUNT 99 a 217 c 147 g 112 t 3 others  
ORIGIN

Query Match 24.9%; Score 35.6; DB 10; Length 578;  
Best Local Similarity 64.6%; Pred. No. 23;  
Matches 53; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 11 CCTCTGCGGCGCTGCTGCTCGGTGGGCGCTCGGACCAAGTCCGACGCGAGA 70  
DB 22 CCTCTCGGCGCTGCGCAGGTGCGGCGCGCGCGGCGAGACGACACGCCATG 81  
QY 71 GCTTGTGCTGCGGCGCGCCACC 92  
DB 82 CCTTACTTTCAGCTGCCACCC 103

RESULT 7  
BF578052

LOCUS BF578052 860 bp mRNA linear EST 12-DEC-2000  
DEFINITION 602094705F1 NCI\_CGAP\_Co24 Mus musculus cDNA clone IMAGE:4209097 5',  
mRNA sequence.

ACCESSION BF578052  
VERSION BF578052.1 GI:11651764  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 860)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabds-r@mail.nih.gov  
Tissue Procurement: Jeffrey E. Green, M.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)





LOCUS	RESULT 10	LOCUS	RESULT 10
CNS0052P	844 bp	DNA	linear
Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR1P16 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.			
AL056652			
AL056652.1	GI:4932342		
GSS.			
Drosophila melanogaster.			
Drosophila melanogaster			
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.			
1 (bases 1 to 844)			
Genoscope.			
Direct Submission			
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : segre@genoscope.cns.fr - Web : www.genoscope.cns.fr)			
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <a href="http://www.fruitfly.org">http://www.fruitfly.org</a> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at <a href="http://bacpac.med.buffalo.edu/drosophila_bac.htm">http://bacpac.med.buffalo.edu/drosophila_bac.htm</a> .			
Location/Qualifiers			
1. 844			
/organism="Drosophila melanogaster"			
/db_xref="taxon:7227"			
/clone="BACR1P16"			
/clone_1ib="RPCI-98"			
/note="end : TET3"			
BASE COUNT	261 a	112 c	92 g 35 t 344 others
ORIGIN			
Query Match	24.5%	Score 35;	DB 17; Length 844;
Best Local Similarity	21.5%;	Pred. No. 38;	
Matches	26;	Conservative 61;	Mismatches 34; Indels 0; Gaps 0;
OY	3	GGCCACTCCCTCTGTGCGCGCTCGCTCGCTCGGTGGGGCTTGGACCAAGTCCGAG	62
Db	490	SMCCSMCCGVSSVCAVCCSSGVSRAVAGCSVGGRVGSSRGRAGRSSGRGGSSVSSGV	549
OY	63	ACGGCAGAGCTGTGCTGTGCGCGCGCCACCGAGCGAGCGGCGCAGAGAGGAGTGGG	122
Db	550	SSSSSVGMGCACSSASSSVSCBSBSSVSSVSSGBVSRGCGRCTVGGVGGGSRVSSCGSS	609
OY	123	C 123	
Db	610	S 610	

BG837739/c		745 bp	mRNA	linear	EST 25-MAY-2001
LOCUS	BG837739				
DEFINITION	Zm10_04h12_A Zm10_AAFC_ECORC_Fusarium graminearum_corn_silk Zea mays cDNA clone Zm10_04h12, mRNA sequence.				
ACCESSION	BG837739				
VERSION	BG837739.1	GI:14204062			
KEYWORDS	EST.				
SOURCE	Zea mays:				
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.				
REFERENCE	Harris,L.J., Balcerzak,M., Allard,S., Saparno,A., Couroux,P., De Moors,A., Hatorri,J.I., Ouellet,T., Robert,L.S., Singh,J.A., Sprott ,D. and Tinker,N.A. Expressed Sequence Tags from Maize Silk Six Hours After Silk Channel Inoculation with Fusarium graminearum Unpublished (2001)				
TITLE	Contact: Harris, Linda J. Eastern Cereal and Oilseed Research Centre Agriculture and Agri-food Canada Bldg. 21, Central Experimental Farm, Ottawa, Ontario, K1A 0C6, CANADA Tel: (613) 759-1314 Fax: (613) 759-6566 Email: harrislj@em.agr.ca.				
JOURNAL	Location/Qualifiers				
COMMENT	1..745 /organism="Zea mays" /cultivar="CO388" /db_xref="taxon:4577" /clone="Zm10_04h12" /clone_lib="Zm10_AAFC_ECORC_Fusarium graminearum_corn_silk"				
FEATURES	/tissue_type="Silk" /dev_stage="4-5 days post-silk emergence" /note="Vector: Bluescript SK+/XhoI-EcoRI, Site 1: EcoRI; Site 2: XhoI; Field-grown corn was silk channel-inoculated in the morning (~10 am) with 1 ml of a macroconidial suspension (500,000 spores/ml) of Fusarium graminearum and silk channels were collected and immediately frozen in liquid nitrogen 6 hours later. RNA was extracted from silk tissue between 1 cm below and above the inoculation point in the silk channel, RNA from five silk channels was pooled."				
BASE COUNT	134 a 211 c 263 g 137 t				
ORIGIN					
Query Match	24.1%; Score 34.4; DB 12; Length 745;				
Best Local Similarity	63.1%; Pred. No. 52;				
Matches	53; Conservative 0; Mismatches 31; Indels 0; Gaps 0;				
OY	56 TCCGCAGACGGCAGACTGTCTGTGCCGGCCCCCGAGCGAGCGCGCGCAGAGG 115   727 TCTACTGACCTGTGACCTCCGCATTCCCAACAGCCCGAGCGAGCTAGCAGCGG 668				
Db	116 GAGTGGCAACTCCATCACTAGGG 139   667 CGGCAGGCACCTCCCTCCTCAAGG 644				
OY					
Db					
RESULT 12					
LOCUS	AG127787	890 bp	DNA	linear	GSS 04-NOV-2001
DEFINITION	Pan troglodytes DNA, clone: PTB-138M02.F, genomic survey sequence.				
ACCESSION	AG127787				
VERSION	AG127787.1	GI:16656952			
KEYWORDS	GSS.				
SOURCE	Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male BAC library clone:PTB-138M02.F. Pan troglodytes Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				

REFERENCE 1 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.  
AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,  
Totoki, Y., Watanabe, H. and Sakaki, Y.  
TITLE BAC end sequences of library PTB  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 890)  
AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,  
Totoki, Y., Watanabe, H. and Sakaki, Y.  
TITLE Direct Submission  
JOURNAL Submitted (02-AUG-2001) Aseo Fujiyama, The Institute of Physical  
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
(E-mail: [chimbesc@sc.riken.go.jp](mailto:chimbesc@sc.riken.go.jp), URL: <http://hgp.gsc.riken.go.jp/>,  
Tel: 81-45-503-9111, Fax: 81-45-503-9170)  
Clones are derived from the chimpanzee BAC library PTB. This BAC end  
was generated during the R&D process and may have higher chance of  
clone tracking errors.  
PRIMERS  
Sequencing: -21M13  
LIBRARY  
Vector : pKS145  
R.Site 1 : SacI  
R.Site 2 : SacI.  
Location/Qualifiers  
1. .890  
/organism="Pan troglodytes"  
/db\_xref="taxon:9598"  
/clone="PTB-138M02.F"  
/sex="male"  
/cell\_type="lymphoblast"  
/clone\_lib="PTB Chimpanzee Male BAC library"  
BASE COUNT 51 a 173 c 478 g 48 t 140 others  
ORIGIN  
Query Match 24.1%; Score 34.4; DB 17; Length 890;  
Best Local Similarity 57.3%; Pred. No. 55;  
Matches 59; Conservative 0; Mismatches 44; Indels 0; Gaps 0;  
QY 18 GCGCGCTCGCTCGCTCGGTGGGGCTCGGACCAAGTCCGACGCGCAGAGCTCTGC 77  
DB 543 GCGCGGGGGCGCGCGCGCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 602  
QY 78 TCTGCGCGCGCCCAACGAGCGAGCGCGCGCAGAGAGGAGTG 120  
DB 603 GCG 645  
RESULT 13  
LOCUS BB613613 623 bp mRNA linear EST 26-OCT-2001  
DEFINITION BB613613 RIKEN full-length enriched, 10 day neonate skin Mus  
ACCESSION BB613613  
VERSION BB613613.1 GI:16454162  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 623)  
AUTHORS Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A.,  
Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda  
, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M.,  
Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki  
, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H.,  
Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toya, T.,  
Muramatsu, M. and Hayashizaki, Y.  
RIKEN Mouse ESTs (Arakawa, T., et al. 2001)  
Unpublished (2001)  
TITLE RIKEN Mouse ESTs (Arakawa, T., et al. 2001)  
JOURNAL Unpublished (2001)  
COMMENT Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: [genome-res@sc.riken.go.jp](mailto:genome-res@sc.riken.go.jp),  
[URL:http://genome.gsc.riken.go.jp/](mailto:URL:http://genome.gsc.riken.go.jp/)  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh  
, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
wagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,  
Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunra  
, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and  
Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000).  
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara  
, Y. and Hayashizaki, Y.  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamana, I., Aizawa  
, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and  
Hayashizaki, Y.  
Computational Analysis of Full-length Mouse cDNAs Compared with  
Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)  
Please visit our web site (<http://genome.gsc.riken.go.jp>) for  
further details.  
e mouse tissues.  
FEATURES  
source  
location/Qualifiers  
1. .623  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="4732496L22"  
/clone\_lib="RIKEN full-length enriched, 10 day neonate  
skin"  
/sex="mixed"  
/tissue\_type="skin"  
/dev\_stage="10 days neonate"  
/lab\_host="DH10B"  
/note="Site 1: SalI; Site 2: BamHI; cDNA library was  
prepared and sequenced in Mouse Genome Encyclopedia  
Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in  
RIKEN. Division of Experimental Animal Research in Riken  
contributed to prepare mouse tissues. 1st strand cDNA was  
primed with a primer [5',  
GAGAGAGAGAGGATCCAGAGCTCTTTTCTTTTCTTTTCTTTVN 3'], cDNA was  
prepared by using trehalose thermo-activated reverse  
transcriptase and subsequently enriched for full-length by  
cap-trapper. cDNA went through one round of normalization  
to Rot = 10.0 and subtraction to Rot = 100.0. Second  
strand cDNA was prepared with the primer adapter of  
sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATCCCCCCCCCCC  
3']. cDNA was cloned into the XhoI and BamHI sites.  
Vector: a modified pBluescript KS(+) after bulk excision  
from Lambda FLC I"  
BASE COUNT 88 a 249 c 177 g 109 t  
ORIGIN  
Query Match 23.8%; Score 34; DB 10; Length 623;  
Best Local Similarity 63.4%; Pred. No. 62;  
Matches 52; Conservative 0; Mismatches 30; Indels 0; Gaps 0;  
QY 11 CCTCTGCGCGCTCGCTCGGTGGGGCTCGGACCAAGTCCGACGCGCAGA 70  
DB 208 CCTCTCGCGCCCTGGCCAGTCTGCGCGCGCGCGCAGAGACACGACGCATG 267  
QY 71 GCTCTGCTCTGCGCGCGCCACC 92









GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 2003, 18:20:14 ; Search time 51.1932 Seconds  
(without alignments)  
1915.734 Million cell updates/sec

Title: US-09-807-802A-13

Perfect score: 3989  
Sequence: 1 MAADGYLPDWLEDNLESGIR.....NNGLYTEPRPIGTRYLTRPL 736

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq.101002.\*  
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3989	100.0	736	21	AAV71167	Adeno-associated v
2	3963	99.3	736	22	AAB59847	AAV6 capsid protei
3	3511	88.0	736	22	AAB59846	AAV3B capsid prote
4	3494	87.6	736	22	AAB59845	AAV3A capsid prote
5	3402.5	85.3	735	22	AAG65792	Adeno-associated v
6	3402.5	85.3	735	22	AAM51508	Adeno-associated v
7	3402.5	85.3	735	22	AAB59844	AAV2 capsid protei
8	3376	84.6	734	22	AAB50326	Adeno-associated v
9	3251	81.5	599	21	AAV71168	Adeno-associated v
10	2906	72.9	534	21	AAV71169	Adeno-associated v

11	2759.5	69.2	598	22	AAG65793	Adeno-associated v
12	2759.5	69.2	598	22	AAM51509	Adeno-associated v
13	2486.5	62.3	734	19	AAW46308	AAV4 VPI capsid pr
14	2481.5	62.2	533	22	AAG65794	Adeno-associated v
15	2477.5	62.1	533	22	AAM51510	Adeno-associated v
16	2220	55.7	724	21	AAV58160	Adeno-associated v
17	2220	55.7	724	23	AAU11405	Adeno-associated v
18	2195.5	55.0	736	18	AAW19000	Duck parvovirus ca
19	2168.5	54.4	732	16	AAR85385	Barbary duck parvo
20	1830.5	45.9	598	19	AAW46313	AAV4 VP2 coat prot
21	1700.5	42.6	588	21	AAV58161	Adeno-associated v
22	1700.5	42.6	588	23	AAU11406	Adeno-associated v
23	1690.5	42.4	544	19	AAW46314	AAV4 VP3 coat prot
24	1676	42.0	534	16	AAR85386	Barbary duck parvo
25	1665	41.7	532	21	AAV58162	Adeno-associated v
26	1665	41.7	532	23	AAU11407	Adeno-associated v
27	1410	35.3	756	21	AAV71231	Capsid protein enc
28	921	23.1	202	23	AAU83481	Adeno-associated v
29	606.5	15.2	781	20	AAV23227	Erythrovirus V9 VP
30	605.5	15.2	781	16	AAW08986	Human parvovirus V
31	490	12.3	554	16	AAW08987	Human parvovirus V
32	490	12.3	554	20	AAV23230	Erythrovirus V9 VP
33	476.5	11.9	543	12	AAR13405	Parvo virus B19 VP
34	425.5	10.7	472	5	AAP40675	Sequence encoded b
35	417.5	10.5	620	22	AAM50111	Feline parvovirus
36	318	8.0	59	23	AAU83456	Parvovirus VPIup c
37	310	7.8	264	12	AAR13407	Parvo virus B19 PA
38	303	7.6	59	23	AAU83454	Parvovirus VPIup c
39	301	7.5	59	23	AAU83453	Parvovirus VPIup c
40	297	7.4	59	23	AAU83452	Parvovirus VPIup c
41	295.5	7.4	370	12	AAR13406	Parvo virus B19 PA
42	282	7.1	686	5	AAP40068	Sequence of a porc
43	262.5	6.6	172	23	AAU83476	Porcine parvovirus
44	255	6.4	59	23	AAU83448	Parvovirus VPIup c
45	252	6.3	59	23	AAU83455	Parvovirus VPIup c

ALIGNMENTS

RESULT 1	
AAV71167	
ID	AAV71167 standard; Protein; 736 AA.
XX	
AC	AAV71167;
XX	
DT	08-SEP-2000 (first entry)
XX	
DE	Adeno-associated virus serotype 1 capsid protein VPI.
XX	
KW	Adeno-associated virus serotype 1; AAV-1; rep protein; capsid protein;
KW	cap protein; recombinant viral vector; gene delivery; gene therapy;
KW	vaccine; transgene; VPI.
XX	
OS	Adeno associated virus serotype 1.
XX	
PN	WO200028061-A2.
XX	
PD	18-MAY-2000.
XX	
PF	02-NOV-1999; 99WO-US25694.
XX	
PR	05-NOV-1998; 98US-0107114.
XX	
PA	(TYPE-) UNIV PENNSYLVANIA.
XX	
PI	Wilson JM, Xiao W;
XX	
DR	WPI; 2000-376571/32.
DR	N-PSDB; AAD00772, AAD00777.
XX	
PT	Novel adeno-associated virus serotype 1 polynucleotide useful for preparation of medicament for delivery of a transgene to a host



XX Claim 7; Page 87-90; 108pp; English.  
PS  
XX  
CC The patent discloses an adeno-associated virus serotype 1 (AAV-1) DNA  
CC which is characterised by two inverted terminal repeats (ITR) and open  
CC reading frames for rep and capsid (cap) proteins. The rep reading frame  
CC encodes four proteins, Rep 78, Rep 68, Rep 52 and Rep 40, while the cap  
CC reading frame encodes three structural proteins, VP1, VP2 and VP3.  
CC The AAV-1 sequence or its fragments particularly ITRs, rep and cap  
CC coding regions, are useful in production of recombinant viral vectors  
CC for gene delivery. These vectors can be used as gene therapy  
CC vectors, vaccine vectors or antisense delivery vectors. The AAV-1 does  
CC not induce the formation of neutralising antibodies specific to any  
CC serotype of AAV hence is useful for transforming host cells, and in the  
CC preparation of a medicament for the delivery of transgene to a host.  
CC The present sequence is an AAV-1 cap protein VP1 which is  
XX useful in the production of recombinant viral vector for gene delivery.  
XX

SQ Sequence 736 AA;  
Query Match 100.0%; Score 3989; DB 21; Length 736;  
Best Local Similarity 100.0%; Pred. No. 8.6e-312;  
Matches 736; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAADGYLPDWLEDNLSGIREWMDLKPAPKPKANQKODDGRGLVLPGYKYLGPFGNGLD 60  
DB 1 MAADGYLPDWLEDNLSGIREWMDLKPAPKPKANQKODDGRGLVLPGYKYLGPFGNGLD 60  
QY 61 KGEFVNADAAALEHDKAYDQOLKAGDNPYLRYNHADAERLQEDTSFGNLTGRAVFQ 120  
DB 61 KGEFVNADAAALEHDKAYDQOLKAGDNPYLRYNHADAERLQEDTSFGNLTGRAVFQ 120  
QY 121 AKKRVLEPLGLVEBGAKTAPGKKRPVEQSPQEPDSSSGIGKTGQOPAKKRLNFGQTGSE 180  
DB 121 AKKRVLEPLGLVEBGAKTAPGKKRPVEQSPQEPDSSSGIGKTGQOPAKKRLNFGQTGSE 180  
QY 181 SVDPDQPLGEPATPAVGPPTMASGGGAPMADNNEGADGVGNASGNWHDSTWLGDRVI 240  
DB 181 SVDPDQPLGEPATPAVGPPTMASGGGAPMADNNEGADGVGNASGNWHDSTWLGDRVI 240  
QY 241 TTSTRTWALPTYNNHLYKQISSASTGASNDNHFGYSTPWGYPDFNRFHCHFSPRDWQRL 300  
DB 241 TTSTRTWALPTYNNHLYKQISSASTGASNDNHFGYSTPWGYPDFNRFHCHFSPRDWQRL 300  
QY 301 INNNWGFPRKRLNFKLENIQVEVTTNDGVTITANLSTVQVFSDEYQIPLYLSAHO 360  
DB 301 INNNWGFPRKRLNFKLENIQVEVTTNDGVTITANLSTVQVFSDEYQIPLYLSAHO 360  
QY 361 GCLPPFPADVFMIPQYGYLTILNNGSQAVGRSSFYCLEYFPPSQMLRTGNFTFSYTFEEVP 420  
DB 361 GCLPPFPADVFMIPQYGYLTILNNGSQAVGRSSFYCLEYFPPSQMLRTGNFTFSYTFEEVP 420  
QY 421 FHSSYAHQSQSLDRMLNPLIDQYLYLNRTQNGSQAQNKDL,FSRGSPPAGMSVQPKNWL 480  
DB 421 FHSSYAHQSQSLDRMLNPLIDQYLYLNRTQNGSQAQNKDL,FSRGSPPAGMSVQPKNWL 480  
QY 481 GPCYRQORVSKTKTDNNNSNFTWTGASKYNNLNGRESIINPGTAMASHKDEDEKFFPMGCV 540  
DB 481 GPCYRQORVSKTKTDNNNSNFTWTGASKYNNLNGRESIINPGTAMASHKDEDEKFFPMGCV 540  
QY 541 MIFGESAGASNTALDNVMTDEBEIKATNPVATERFGTVA,VNFQSSSTDPATGDVHAMG 600  
DB 541 MIFGESAGASNTALDNVMTDEBEIKATNPVATERFGTVA,VNFQSSSTDPATGDVHAMG 600  
QY 601 ALPGMWQDRDVLQGPIMAKIPHTDGHFHPSPLMGGFGLKNPPQOLIKNTVPANPPA 660  
DB 601 ALPGMWQDRDVLQGPIMAKIPHTDGHFHPSPLMGGFGLKNPPQOLIKNTVPANPPA 660  
QY 661 EFSATKFAFITQYSTGOVSVEIEWELOKENSKRWNPEVQYTSNYAKSANVDFTVDNGL 720  
DB 661 EFSATKFAFITQYSTGOVSVEIEWELOKENSKRWNPEVQYTSNYAKSANVDFTVDNGL 720  
QY 721 YTEPRPIGTRYLTRPL 736

DB 721 YTEPRPIGTRYLTRPL 736  
RESULT 2  
AAB59847  
ID AAB59847 standard; Protein; 736 AA.  
XX  
AC AAB59847;  
XX  
DT 28-MAR-2001 (first entry)  
XX  
DE AAV6 capsid protein VP1.  
XX  
KW AAV6; gene therapy; AAV viral vector; cystic fibrosis; cancer; AIDS;  
KW atherosclerosis; sickle cell anaemia; thalassemia;  
KW blood clotting disorder; diabetes; capsid protein VP1.  
XX  
OS Adeno associated virus.  
XX  
XX  
PN US6156303-A.  
XX  
PD 05-DEC-2000.  
XX  
PF 11-JUN-1997; 97US-0873168.  
XX  
PR 11-JUN-1997; 97US-0873168.  
XX  
PA (UNIW ) UNIV WASHINGTON.  
PI Russell DW, Rutledge EA;  
XX  
DR WPI; 2001-060164/07.  
XX

PT Adeno-associated virus serotype 6 and viral vector derived from it for  
PT gene therapy of cystic fibrosis, cancer, acquired immunodeficiency  
PT syndrome, sickle cell anemia, thalassemia and diabetes  
XX  
PS Disclosure; Fig 2; 50pp; English.  
XX

CC The present invention relates to adeno-associated virus serotypes. The  
CC present sequence is capsid protein VP1 of one such serotype (AAV6). AAV6  
CC can be used to construct AAV viral vectors for use in gene therapy for a  
CC range of disorders: cystic fibrosis, cancer, AIDS, atherosclerosis,  
CC sickle cell anaemia, thalassemia, blood clotting disorders and diabetes.  
CC The AAV viral vectors have increased transduction efficiency of a  
CC particular host cell as the AAV virion containing the AAV vector genome  
CC can be modified to express a capsid protein of an AAV serotype that  
CC transduces the selected host cell.  
XX  
SQ Sequence 736 AA;

Query Match 99.3%; Score 3963; DB 22; Length 736;  
Best Local Similarity 99.2%; Pred. No. 1.1e-309;  
Matches 730; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 MAADGYLPDWLEDNLSGIREWMDLKPAPKPKANQKODDGRGLVLPGYKYLGPFGNGLD 60  
DB 1 MAADGYLPDWLEDNLSGIREWMDLKPAPKPKANQKODDGRGLVLPGYKYLGPFGNGLD 60  
QY 61 KGEFVNADAAALEHDKAYDQOLKAGDNPYLRYNHADAERLQEDTSFGNLTGRAVFQ 120  
DB 61 KGEFVNADAAALEHDKAYDQOLKAGDNPYLRYNHADAERLQEDTSFGNLTGRAVFQ 120  
QY 121 AKKRVLEPLGLVEBGAKTAPGKKRPVEQSPQEPDSSSGIGKTGQOPAKKRLNFGQTGSE 180  
DB 121 AKKRVLEPLGLVEBGAKTAPGKKRPVEQSPQEPDSSSGIGKTGQOPAKKRLNFGQTGSE 180  
QY 181 SVDPDQPLGEPATPAVGPPTMASGGGAPMADNNEGADGVGNASGNWHDSTWLGDRVI 240  
DB 181 SVDPDQPLGEPATPAVGPPTMASGGGAPMADNNEGADGVGNASGNWHDSTWLGDRVI 240  
QY 241 TTSTRTWALPTYNNHLYKQISSASTGASNDNHFGYSTPWGYPDFNRFHCHFSPRDWQRL 300

|||||  
Db 241 TTSTRTALPTYNNHLYKQISSASTGASNDNHFGYSTPWGTFDNRHCHFSPRDWRL 300  
QY 301 INNMGFRPKRLNFKLFNIQVKEVTTNDGVTTIANNLTSTVQVFSDEYQLPYVLGSAHQ 360  
Db 301 INNMGFRPKRLNFKLFNIQVKEVTTNDGVTTIANNLTSTVQVFSDEYQLPYVLGSAHQ 360  
QY 361 GCLPPPADVFMIPQYGYLTLNNGSAVGRSSFYCLEYFPSPQMLRTGNNFTFSYTFEEVP 420  
Db 361 GCLPPPADVFMIPQYGYLTLNNGSAVGRSSFYCLEYFPSPQMLRTGNNFTFSYTFEEVP 420  
QY 421 FHSSYAHQSOLDRLMNPILIDQYLYLNRTQNGSGAQNKDILFSRGSFAGMSVQPKMWLP 480  
Db 421 FHSSYAHQSOLDRLMNPILIDQYLYLNRTQNGSGAQNKDILFSRGSFAGMSVQPKMWLP 480  
QY 481 GPCYRQQRVSKTKTDNNSNFTWTGASKYNLNGRESIINPGTAMASHKODEDKFFPMGCV 540  
Db 481 GPCYRQQRVSKTKTDNNSNFTWTGASKYNLNGRESIINPGTAMASHKODEDKFFPMGCV 540  
QY 541 MIFGKESAGASNTALDNVMTDDEEIKATNPVATERFGTVAVNFQSSSTDPAATGVDHAMG 600  
Db 541 MIFGKESAGASNTALDNVMTDDEEIKATNPVATERFGTVAVNFQSSSTDPAATGVDHAMG 600  
QY 601 ALPGMWQDRDVLQGPFWAKIPHTDGHFHPSPPLMGFGGLKNPPEQILIKNTPVPANPPA 660  
Db 601 ALPGMWQDRDVLQGPFWAKIPHTDGHFHPSPPLMGFGGLKNPPEQILIKNTPVPANPPA 660  
QY 661 EFSATKRFASFITQYSTGQVSVEIEMELQKENSKRWNPEVQYTSNYAKSANVDFTVDNNG 720  
Db 661 EFSATKRFASFITQYSTGQVSVEIEMELQKENSKRWNPEVQYTSNYAKSANVDFTVDNNG 720  
QY 721 YTEPRPIGTRYLTRPL 736  
Db 721 YTEPRPIGTRYLTRPL 736

RESULT 3  
AAB59846 ID AAB59846 standard; Protein; 736 AA.  
XX AC AAB59846;  
XX DT 28-MAR-2001 (first entry)  
XX DE AAV3B capsid protein VP1.  
XX KW AAV3B; gene therapy; AAV viral vector; cystic fibrosis; cancer; AIDS;  
KW atherosclerosis; sickle cell anaemia; thalassaemia;  
KW blood clotting disorder; diabetes; capsid protein VP1.  
XX OS Adeno associated virus.  
XX PN US6156303-A.  
XX PD 05-DEC-2000.  
XX PF 11-JUN-1997; 97US-0873168.  
XX PR 11-JUN-1997; 97US-0873168.  
XX PA (UNITW ) UNIV WASHINGTON.  
XX PI Russell DW, Rutledge EA;  
XX DR WPI; 2001-060164/07.  
XX PT Adeno-associated virus serotype 6 and viral vector derived from it for  
PT gene therapy of cystic fibrosis, cancer, acquired immunodeficiency  
PT syndrome, sickle cell anemia, thalassemia and diabetes  
XX PS Disclosure; Fig 2; 50pp; English.  
XX CC The present invention relates to adeno-associated virus serotypes. The

CC present sequence is capsid protein VP1 of one such serotype (AAV3B).  
CC AAV3B can be used to construct AAV viral vectors for use in gene therapy  
CC for a range of disorders: cystic fibrosis, cancer, AIDS, atherosclerosis,  
CC sickle cell anaemia, thalassaemia, blood clotting disorders and diabetes.  
CC The AAV viral vectors have increased transduction efficiency of a  
CC particular host cell as the AAV virion containing the AAV vector genome  
CC can be modified to express a capsid protein of an AAV serotype that  
CC transduces the selected host cell.  
SQ Sequence 736 AA;

Query Match 88.0%; Score 3511; DB 22; Length 736;  
Best Local Similarity 86.8%; Pred. No. 2.4e-273;  
Matches 640; Conservative 38; Mismatches 57; Indels 2; Gaps 2;

QY 1 MAADGYLPDWLEDNLSEGIREWMDLKPGAPKPKANQOKODDGRGLVLPGYKYLGPENGLD 60  
Db 1 MAADGYLPDWLEDNLSEGIREWMDLKPGVPQPKANQOHQDNRGLVLPGYKYLGPENGLD 60  
QY 61 KGEPVNAADAALAEHDKAYDQOLKAGDNPYLRYNHADAERLQEDTSFGGNLGRAVFQ 120  
Db 61 KGEPVNEADAALAEHDKAYDQOLKAGDNPYLRYNHADAERLQEDTSFGGNLGRAVFQ 120  
QY 121 AKKRVLEPLGLVEEGAKTAPGKKRPVEQSPOEPDSSSGIGTKQOPAKKRLNFGQTGSE 180  
Db 121 AKKRVLEPLGLVEEGAKTAPGKKRPVDSPOEPDSSSGVGSKGQPARKRLNFGQTGSE 180  
QY 181 SVDPDQPLGEPAPPAVPTMASGGAPMADNNEGADGVNAGNMHCDSTWLGDRVI 240  
Db 181 SVDPDQPLGEPAPPAVPTSLGSNTMASGGAPMADNNEGADGVNAGNMHCDSTWLGDRVI 240  
QY 241 TTSTRTALPTYNNHLYKQISSASTGASNDNHFGYSTPWGTFDNRHCHFSPRDWRL 300  
Db 241 TTSTRTALPTYNNHLYKQISSQS-GASNDNHFGYSTPWGTFDNRHCHFSPRDWRL 299  
QY 301 INNMGFRPKRLNFKLFNIQVKEVTTNDGVTTIANNLTSTVQVFSDEYQLPYVLGSAHQ 360  
Db 300 INNMGFRPKRLNFKLFNIQVKEVTQNDGVTTIANNLTSTVQVFTDSEYQLPYVLGSAHQ 359  
QY 361 GCLPPPADVFMIPQYGYLTLNNGSAVGRSSFYCLEYFPSPQMLRTGNNFTFSYTFEEVP 420  
Db 360 GCLPPPADVFMIPQYGYLTLNNGSAVGRSSFYCLEYFPSPQMLRTGNNFTFSYTFEEVP 419  
QY 421 FHSSYAHQSOLDRLMNPILIDQYLYLNRTQ-NQSGAQNKDILFSRGSFAGMSVQPKMWLP 479  
Db 420 FHSSYAHQSOLDRLMNPILIDQYLYLNRTQGTSTGTTNQSRLLFSQAGPQMSLQARWML 479  
QY 480 PGPCYRQQRVSKTKTDNNSNFTWTGASKYNLNGRESIINPGTAMASHKODEDKFFPMGCV 539  
Db 480 PGPCYRQQRVSKTKTDNNSNFTWTGASKYNLNGRDSLVPNGPAMASHKODEDKFFPMGCV 539  
QY 540 VMIFGKESAGASNTALDNVMTDDEEIKATNPVATERFGTVAVNFQSSSTDPAATGVDHAM 599  
Db 540 NLIFGKEGTTASNAELDNVMTDDEEIRTTNPVATERFGTVAVNNLQSSNTAFTTRIVNDQ 599  
QY 600 GALPGMWQDRDVLQGPFWAKIPHTDGHFHPSPPLMGFGGLKNPPEQILIKNTPVPANPP 659  
Db 600 GALPGMWQDRDVLQGPFWAKIPHTDGHFHPSPPLMGFGGLKNPPEQIMIKNTPVPANPP 659  
QY 660 AFSATKRFASFITQYSTGQVSVEIEMELQKENSKRWNPEVQYTSNYAKSANVDFTVDNNG 719  
Db 660 TTFSPAKRFASFITQYSTGQVSVEIEMELQKENSKRWNPEIQYTSNYKNSVNVDFTVDNNG 719  
QY 720 LYTEPRPIGTRYLTRPL 736  
Db 720 VYSEPRPIGTRYLTRPL 736

RESULT 4  
AAB59845 ID AAB59845 standard; Protein; 736 AA.  
XX AC AAB59845;

XX 28-MAR-2001 (first entry)  
DT AAV3A capsid protein VP1.  
XX  
XX AAV3A; gene therapy; AAV viral vector; cystic fibrosis; cancer; AIDS;  
KM atherosclerosis; sickle cell anaemia; thalassaemia;  
KW blood clotting disorder; diabetes; capsid protein VP1.  
XX  
OS Adeno associated virus.  
XX  
PN US6156303-A.  
XX  
PD 05-DEC-2000.  
XX  
XX 11-JUN-1997; 97US-0873168.  
PF 11-JUN-1997; 97US-0873168.  
XX  
PR 11-JUN-1997; 97US-0873168.  
XX  
PA (UNIW ) UNIV WASHINGTON.  
XX  
PI Russell DW, Rutledge EA;  
XX  
XX WPI; 2001-060164/07.  
DR  
XX Adeno-associated virus serotype 6 and viral vector derived from it for  
PT gene therapy of cystic fibrosis, cancer, acquired immunodeficiency  
PT syndrome, sickle cell anemia, thalassemia and diabetes  
XX  
XX  
PS Disclosure; Fig 2; 50pp; English.  
XX  
CC The present invention relates to adeno-associated virus serotypes. The  
CC present sequence is capsid protein VP1 of one such serotype (AAV3A).  
CC AAV3A can be used to construct AAV viral vectors for use in gene therapy  
CC for a range of disorders: cystic fibrosis, cancer, AIDS, atherosclerosis,  
CC sickle cell anaemia, thalassaemia, blood clotting disorders and diabetes.  
CC The AAV viral vectors have increased transduction efficiency of a  
CC particular host cell as the AAV virion containing the AAV vector genome  
CC can be modified to express a capsid protein of an AAV serotype that  
CC transduces the selected host cell.  
XX  
SQ Sequence 736 AA;  
  
Query Match 87.6%; Score 3494; DB 22; Length 736;  
Best Local Similarity 86.4%; Pred. No. 5.6e-272;  
Matches 637; Conservative 39; Mismatches 59; Indels 2; Gaps 2;  
  
QY 1 MAADGYLPDWLEDNLSEGIREWMDLKGAPKPKANQQKODDGRGLVLPGYKYLGPFGNLD 60  
DB 1 MAADGYLPDWLEDNLSEGIREWMDLKGAPKPKANQQKODDGRGLVLPGYKYLGPFGNLD 60  
QY 61 KGEPVNAADAALAEHDKAYDQQLKAGDNPYLRYNHADAEFOERLQEDTSFGNLTGRAVFQ 120  
DB 61 KGEPVNAADAALAEHDKAYDQQLKAGDNPYLRYNHADAEFOERLQEDTSFGNLTGRAVFQ 120  
QY 121 AKKRVLEPLGLVEGAKTAPGKKRPVEQSPQEPDSSSGIGKTGQOPAKKRLNFGQTDSE 180  
DB 121 AKKRVLEPLGLVEGAKTAPGKKRPVEQSPQEPDSSSGIGKTGQOPAKKRLNFGQTDSE 180  
QY 181 SVDPDQPLGEPAPTPAVGPTTMASGGGAPVADNNEGADGVGNASGWHCDSTWLGDRVI 240  
DB 181 SVDPDQPLGEPAPTPAVGPTTMASGGGAPVADNNEGADGVGNASGWHCDSTWLGDRVI 240  
QY 241 TTSTRTVALPTYNHLYKQISSASTGASNDNHYFGYSTPWGTFDFNRFHGFSPRDWRL 300  
DB 241 TTSTRTVALPTYNHLYKQISSASTGASNDNHYFGYSTPWGTFDFNRFHGFSPRDWRL 300  
QY 301 INNNWGFRPKRLNFKLFNIQVEKVTNDGVTTIANLSTVQVFSDEYQLPYVLGSAHQ 360  
DB 301 INNNWGFRPKRLNFKLFNIQVEKVTNDGVTTIANLSTVQVFSDEYQLPYVLGSAHQ 360  
QY 361 GCLPPFPADVFMIDQGYLLTLNNGSQAVGRSSFYCLEYFSPQMLRTGNFTFSYTFEEVP 420  
DB 361 GCLPPFPADVFMIDQGYLLTLNNGSQAVGRSSFYCLEYFSPQMLRTGNFTFSYTFEEVP 420

DB 360 GCLPPFPADVFMIDQGYLLTLNNGSQAVGRSSFYCLEYFSPQMLRTGNFTFSYTFEEVP 419  
QY 421 FHSSYAHSQSLDRLMNPPLIDQYLYLNRTO-NOSGSAQNKDLFSRGSFAGMSVQPKMWL 479  
DB 420 FHSSYAHSQSLDRLMNPPLIDQYLYLNRTO-NOSGSAQNKDLFSRGSFAGMSVQPKMWL 479  
QY 480 PGPCYRQQRVSKTKTDNNNSNFTWTSKYNLNGRESIINPGTAMASHKDEDEKFFPMG 539  
DB 480 PGPCYRQQRVSKTKTDNNNSNFTWTSKYNLNGRESIINPGTAMASHKDEDEKFFPMG 539  
QY 540 VMIPEKESAGASNTALDNVMTDDEEIKATNPVATERFETVAVNFQSSSTDPAFGVHAM 599  
DB 540 VMIPEKESAGASNTALDNVMTDDEEIKATNPVATERFETVAVNFQSSSTDPAFGVHAM 599  
QY 600 GALPGMWQDRDVLQGPPIWAKIPHTDGHFHPSPLMGFGILKNPPQILIKNTPVANPP 659  
DB 600 GALPGMWQDRDVLQGPPIWAKIPHTDGHFHPSPLMGFGILKNPPQILIKNTPVANPP 659  
QY 660 AEFSAITKFAFITQYSTGVSEIEMELQKENSKRWNPEVOYTSNYAKSANVDFVDNNG 719  
DB 660 AEFSAITKFAFITQYSTGVSEIEMELQKENSKRWNPEVOYTSNYAKSANVDFVDNNG 719  
QY 720 LYTEPRPIGTRYLTRPL 736  
DB 720 LYTEPRPIGTRYLTRPL 736  
  
RESULT 5  
AAG65792  
AAG65792 standard; Protein; 735 AA.  
XX  
AC AAG65792;  
XX  
DT 11-FEB-2002 (first entry)  
XX  
DE Adeno-associated virus 2 (AAV-2) major coat protein VP1.  
XX  
KW Recombinant viral vector; RVV; capsid; parvovirus; transgene; cytosratic;  
KW inverted terminal repeat; noctropic; neuroprotective; antianemic; ITR;  
KW antidiabetic; antitumour; gene therapy; adeno-associated virus; AAV;  
KW major coat protein; AAV-2; VP1.  
XX  
XX Adeno-associated virus 2.  
OS  
XX WO200168888-A2.  
PN  
XX 20-SEP-2001.  
PD  
XX 13-MAR-2001; 2001WO-US07927.  
PF  
XX 14-MAR-2000; 2000US-189110P.  
PR  
XX (NEUR-) NEUROLOGIX INC.  
PA  
PI Xiao W, During MJ;  
XX  
XX WPI; 2001-596912/67.  
DR  
XX N-PSDB; AAI66974.  
DR  
XX  
PT Recombinant viral vector useful in improving gene therapy in a subject,  
PT and for increasing efficiency of entry into a cell, comprises a  
PT chimeric capsid having one non-native amino acid sequence and a desired  
PT transgene  
XX  
XX  
PS Disclosure; Page 50; 53pp; English.  
XX  
CC The invention provides a recombinant viral vector (RVV) comprising a  
CC chimeric capsid (I) having at least one non-native amino acid sequence,  
CC derived from a capsid protein domain of parvovirus (II), a virus (III),  
CC or their combination, and a transgene flanked 5' and 3' by inverted  
CC terminal repeat (ITR) sequences, derived from (II), (III), or their  
CC combination. The RVV is useful for improving gene therapy in a subject  
CC with a disorder, and for increasing the efficiency of entry into a cell,



CC which involves producing (I) encapsulating a viral vector, and contacting  
 CC a cell with RVV having (I) such that (I) binds to an attachment site on  
 CC the cell surface and permits the vector to enter the cell efficiently.  
 CC A pharmaceutical composition comprising RVV with (I) containing a  
 CC transgene sequence associated with a disease or a disorder such that  
 CC expression of the transgene would result in amelioration of the disease  
 CC or disorder such as inherited neurological and metabolic diseases e.g.  
 CC lysosomal storage disease, Lesch-Nyhan syndrome, amyloid polynuropathy,  
 CC Alzheimer amyloid, Duchenne's muscular dystrophy, diseases of the blood  
 CC e.g. sickle-cell anemia, clotting disorders and thalassemias, cystic  
 CC fibrosis, diabetes, diseases associated with hormone deficiencies,  
 CC retinoblastoma and various types of neoplastic cells which include  
 CC tumours especially central nervous system tumours, neoplasms, carcinomas,  
 CC sarcomas, leukemias and lymphoma. The present sequence represents the  
 CC adeno-associated virus 2 (AAV-2) major coat protein VP1. AAV-2 sequences  
 CC are used in the construction of a chimeric vector.

XX  
 XX Sequence 735 AA;

Query Match 85.3%; Score 3402.5; DB 22; Length 735;  
 Best Local Similarity 83.3%; Pred. No. 1.3e-264;  
 Matches 613; Conservative 51; Mismatches 71; Indels 1; Gaps 1;

QY 1 MAADGYLPDWLENDLSEGIKREWWDLKPGAPKPKANQKQDDGRGLVLPYKYLGPFGNLD 60  
 DB 1 MAADGYLPDWLEDTLSEGIKRWKWKLPKPPPKPAERHKDSDRGVLPGYKYLGPFGNLD 60  
 QY 61 KGEPVNAADAALAEHDKAYDQOLKAGDNPYLRYNHADAERFQERLOEDTSFGNLTGRAVFQ 120  
 DB 61 KGEPVNEADAALAEHDKAYDRQLDSGDNLYLKYNHADAERFQERLKEDTSFGNLTGRAVFQ 120  
 QY 121 AKKRVLEPLGLVEEGAKTAPGKKRPVEQSPQEPDSSSGIGKTGQOPAKKRLNFGQTDSE 180  
 DB 121 AKKRVLEPLGLVEEPVKTAPEGKRPVEHSVPEPDSSSGTGKAGQOPAKKRLNFGQTDAD 180  
 QY 181 SVPDPQPLGEPPTPAVGPPTMASGGGAPMADNNEGADGVGNASGNWHCDSTWLGDRVI 240  
 DB 181 SVPDPQPLGQPPAPSGLGTNTMATGSGAPMADNNEGADGVGNSSGNWHCDSTWMDRVI 240  
 QY 241 TTSTRTWALLPTYNHLYKQISSASTGASNDNHFGYSTPWGYPDENRHHCHFSPRDWQRL 300  
 DB 241 TTSTRTWALLPTYNHLYKQISSQS-GASNDNHFGYSTPWGYPDENRHHCHFSPRDWQRL 299  
 QY 301 INNNGWFRPKRLNFKLFNIQVKEVTNDGVTTIANNLTSTVQVPSDSEYQLPYVLGSAHQ 360  
 DB 300 INNNGWFRPKRLNFKLFNIQVKEVTNDGVTTIANNLTSTVQVFTDSEYQLPYVLGSAHQ 359  
 QY 361 GCLPPFPADYEMIPQYGYLTINNNGSQAVGRSSFYCLEYFSPQMLRTGNNFTFSYTFEEVP 420  
 DB 360 GCLPPFPADYEMVWPQYGYLTINNNGSQAVGRSSFYCLEYFSPQMLRTGNNFTFSYTFEDVP 419  
 QY 421 FHSSYAHSGSLDRMLNPLIDQYLYLNRTQONQSGSAQNKDLFSGRSPAGMSVQPKNWL 480  
 DB 420 FHSSYAHSGSLDRMLNPLIDQYLYLSRTNTPSGTTQSRLOFSQAGASDIRDQSRNWL 479  
 QY 481 GPCYRQQRVSKTKTDNNNSNFTWTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMGCV 540  
 DB 480 GPCYRQQRVSKTSADNNNSSEYSGTATKYLHNGRDSLVPNPGPAMASHKDEDEKFFPQSGV 539  
 QY 541 MIFGESAGASNTALDNVMTDEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMG 600  
 DB 540 LIFGKGSEKTNVDIEKVMITDEEIRFTNPVATEQYGSVSTNLQKGNRQAATADAVNTQG 599  
 QY 601 ALPGMWODRDVYLQGPWAKIPHTDGHFSPPLMGFGGLKNPPOILLKNTVPANPAP 660  
 DB 600 VLPGMWODRDVYLQGPWAKIPHTDGHFSPPLMGFGGLKNPPOILLKNTVPANPAPST 659  
 QY 661 EFSATKASFITQYSTGQVSVEIEMELQKENSKRWNPEVQYTSNYAKSANDFTVDNNG 720  
 DB 660 TFSAAKASFITQYSTGQVSVEIEMELQKENSKRWNPEIQYTSNYAKSVNVDFTVDINGV 719  
 QY 721 YTEPRPIGTRYLTRPL 736  
 DB 721 YTEPRPIGTRYLTRPL 736

DB 720 YSEPRPIGTRYLTRNL 735

# RESULT 6

AAM51508

ID AAM51508 standard; protein; 735 AA.

XX AAM51508;

DT 02-JAN-2002 (first entry)

XX Adeno-associated virus VP1 capsid protein.

XX Adeno-associated virus; AAV; VP1; capsid; virus-like particle;  
 KW nuclear localisation signal; VP3.

OS Adeno associated virus.

PN JP2001169777-A.

PD 26-JUN-2001.

PF 30-JUL-1999; 99JP-0249140.

PR 30-JUL-1999; 99JP-0249140.

PA (HAND/) HANDA H.

DR WPI; 2001-599854/68.

PT New virus-like particles from VP3 capsid protein of adeno-associated  
 PT virus, comprise a peptide containing a nucleus-shifting signal  
 PT connected to its N-terminal

PS Disclosure; Page 10-13; 33pp; Japanese.

CC The present sequence is provided in a specification relating to a  
 CC virus-like particle-forming protein, and to a peptide containing a  
 CC nuclear-shifting signal at its N-terminus that can form virus-like  
 CC particles by shifting to the nucleus of the animal cell in which it  
 CC is expressed. The method is used for forming virus-like particles  
 CC from the VP3 capsid protein of adeno-associated virus (AAV).

XX Sequence 735 AA;

Query Match 85.3%; Score 3402.5; DB 22; Length 735;  
 Best Local Similarity 83.3%; Pred. No. 1.3e-264;  
 Matches 613; Conservative 51; Mismatches 71; Indels 1; Gaps 1;

QY 1 MAADGYLPDWLENDLSEGIKREWWDLKPGAPKPKANQKQDDGRGLVLPYKYLGPFGNLD 60  
 DB 1 MAADGYLPDWLEDTLSEGIKRWKWKLPKPPPKPAERHKDSDRGVLPGYKYLGPFGNLD 60  
 QY 61 KGEPVNAADAALAEHDKAYDQOLKAGDNPYLRYNHADAERFQERLOEDTSFGNLTGRAVFQ 120  
 DB 61 KGEPVNEADAALAEHDKAYDRQLDSGDNLYLKYNHADAERFQERLKEDTSFGNLTGRAVFQ 120  
 QY 121 AKKRVLEPLGLVEEGAKTAPGKKRPVEQSPQEPDSSSGIGKTGQOPAKKRLNFGQTDSE 180  
 DB 121 AKKRVLEPLGLVEEPVKTAPEGKRPVEHSVPEPDSSSGTGKAGQOPAKKRLNFGQTDAD 180  
 QY 181 SVPDPQPLGEPPTPAVGPPTMASGGGAPMADNNEGADGVGNASGNWHCDSTWLGDRVI 240  
 DB 181 SVPDPQPLGQPPAPSGLGTNTMATGSGAPMADNNEGADGVGNSSGNWHCDSTWMDRVI 240  
 QY 241 TTSTRTWALLPTYNHLYKQISSASTGASNDNHFGYSTPWGYPDENRHHCHFSPRDWQRL 300  
 DB 241 TTSTRTWALLPTYNHLYKQISSQS-GASNDNHFGYSTPWGYPDENRHHCHFSPRDWQRL 299  
 QY 301 INNNGWFRPKRLNFKLFNIQVKEVTNDGVTTIANNLTSTVQVPSDSEYQLPYVLGSAHQ 360  
 DB 300 INNNGWFRPKRLNFKLFNIQVKEVTNDGVTTIANNLTSTVQVFTDSEYQLPYVLGSAHQ 359

QY	361	GCLEPPFPADVEMIPOYGYLLTNNGSQAVGSSSFYCLEYFPSQMLRTGNNFTSYTFEEVP	420
Db	360	GCLEPPFPADVEMVPOGYLLTNNGSQAVGSSSFYCLEYFPSQMLRTGNNFTSYTFEEDVP	419
QY	421	FHSSYAHSQSLDRLMNPOLIDOYLYYLNRTQNGSGSAQONKDLJFSRGSFAGMSVQPKNWL	480
Db	420	FHSSYAHSQSLDRLMNPOLIDOYLYYLNRTQNGSGTTQSRLOFSQAGASDIRDQSRNWL	479
QY	481	GPCYRQQRVSKTKTDNNNSNFTWTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMGCV	540
Db	480	GPCYRQQRVSKTSADNNNSSEYSWTGATKYHLNGRDSLVPNGPAMASHKDEEKFFPQSGV	539
QY	541	MIFGKESAGASNTALDNVMITDEEIEIKATNPVATERFGTVAVNFQSSSTDPATGVDHAMG	600
Db	540	LIFGKGSEKTNVDIEKVMITDEEIRRTNPVATEQYGSVSTNLQGRNGRQATADVNTQG	599
QY	601	ALPGWVQDRDYYLQGPWAKIPHDTGHFHPSPLMGGFGLKNPPQILLKNTPEVPANPRA	660
Db	600	VLPBWVQDRDYYLQGPWAKIPHDTGHFHPSPLMGGFGLKNPPQILLKNTPEVPANPST	659
QY	661	EFSATKFASFITQYSTGQVSVEIEMELQKENSKRWNPEVOYTSNYAKSANVDFTVDNNGL	720
Db	660	TFSAAKFASFITQYSTGQVSVEIEMELQKENSKRWNPEIOYTSNYAKSANVDFTVDITNGV	719
QY	721	YTBPRPIGTRYLTRPL	736
Db	720	YSEPRPIGTRYLTRNL	735

RESULT 7	
AAB59844	
ID	AAB59844 standard; Protein; 735 AA.
XX	
AC	AAB59844;
XX	
DT	28-MAR-2001 (first entry)
XX	
DE	AAV2 capsid protein VP1.
XX	
KM	AAV2; gene therapy; AAV viral vector; cystic fibrosis; cancer; AIDS;
KW	atherosclerosis; sickle cell anaemia; thalassaemia;
KX	blood clotting disorder; diabetes; capsid protein VP1.
XX	
OS	Adeno associated virus.
XX	
PN	US6156303-A.
XX	
PD	05-DEC-2000.
XX	
PF	11-JUN-1997; 97US-0873168.
XX	
PR	11-JUN-1997; 97US-0873168.
XX	
PA	(UNIW ) UNIV WASHINGTON.
XX	
PI	Russell DW, Rutledge EA;
XX	
DR	WPI; 2001-060164/07.
XX	
PT	Adeno-associated virus serotype 6 and viral vector derived from it for
PT	gene therapy of cystic fibrosis, cancer, acquired immunodeficiency
PT	syndrome, sickle cell anemia, thalassemia and diabetes
XX	
PS	Claim 7; Fig 2; 50pp; English.
XX	
CC	The present invention relates to adeno-associated virus serotypes. The
CC	present sequence is capsid protein VP1 of one such serotype (AAV2). AAV2
CC	can be used to construct AAV viral vectors for use in gene therapy for a
CC	range of disorders: cystic fibrosis, cancer, AIDS, atherosclerosis,
CC	sickle cell anaemia, thalassaemia, blood clotting disorders and diabetes.
CC	The AAV viral vectors have increased transduction efficiency of a
CC	particular host cell as the AAV virion containing the AAV vector genome
CC	can be modified to express a capsid protein of an AAV serotype that

CC	transduces the selected host cell.
XX	Sequence 735 AA;
XX	Query Match 85.3%; Score 3402.5; DB 22; Length 735;
	Best Local Similarity 83.3%; Pred. No. 1.3e-264;
	Matches 613; Conservative 51; Mismatches 71; Indels 1; Gaps 1
QY	1 MAADGYLPDWLEDNLSGEGIREWMDLKPGAPKPKANQOKDDGRGLVPGYKYLGPFGNGLD 60
Db	1 MAADGYLPDWLEDTLSEGIQWQWKLKPGPPPKPAERHKDSDRGVLPGYKYLGPFGNGLD 60
QY	61 KGEVNAADAALAEHDKAYDQOLKAGDNPYLRYNHADAEEFERLQEDTSFGNLRGAVFQ 120
Db	61 KGEVNAADAALAEHDKAYRQRLSDGDNPYLKYNHADAEEFERLQEDTSFGNLRGAVFQ 120
QY	121 AKKRVLEPLGLVEEGAKTAPGKKRPVEQSPQEPDSSSGIGTGOOPAKKRLNFGQTDSE 180
Db	121 AKKRVLEPLGLVEEPVKTAAGKKRPVEHSPVEPDSSSGTGKAGQOPARKRLNFGQTDAD 180
QY	181 SVDPDQPLGEPPTPAVGPPTMASGGGAPMADNNEGADGVGNASGNWHCDSTWLGDRV 240
Db	181 SVDPDQPLGQPPAAPSGLGNTMATGSGAPMADNNEGADGVGNSSGNWHCDSTWMDRVI 240
QY	241 TTSTRTWALPTYNNHLKYQSSASTGASNDNHFGYSTPWGTFDENRFCHFSPRDWQRL 300
Db	241 TTSTRTWALPTYNNHLKYQSSQS - GASNDNHFGYSTPWGTFDENRFCHFSPRDWQRL 299
QY	301 INNWGFRPKRLNFKLFNIQVKEVTINDGVTIANNLTSTVOVFSDEYQLPYVLGSAHQ 360
Db	300 INNWGFRPKRLNFKLFNIQVKEVTQNDGTTIANNLTSTVOYFTDSEYQLPYVLGSAHQ 359
QY	361 GCLPFPFADVFMIPOYGYLTLLNNGSQAVGRSSFYCLEYFPQOMLRTGNNFTFSYTFEEVP 420
Db	360 GCLPFPFADVFVMPQYGYLTLLNNGSQAVGRSSFYCLEYFPQOMLRTGNNFTFSYTFEDVP 419
QY	421 FHSSYAHQSOLDRLMNPILDOYLYLNRTONQSGSAQNKLLFSRGSFAGMSVQPKNWL 480
Db	420 FHSSYAHQSOLDRLMNPILDOYLYLSRTNTPSGTTQSRLOFSQAGASDIRDQSRNWL 479
QY	481 GPCYRQORVSKTKTDNNNSNFTWTGASKYNLNGRESINPGTAMASHKDEDEKFFPMSCV 540
Db	480 GPCYRQORVSKTSADNNNSSEYWTGATKYHLNGRDSLVPGPAMASHKDEDEKFFPQSGV 539
QY	541 MIFGESAGASNTALDWNMTDEEIKATNPVATERFGTAVANFQSSSTDPATGDVHAMG 600
Db	540 LIFGKQSEKTNVDIEKVMITDEEIRTTNPVATEQYGSVSTNLQGRNQATADAVNTQG 599
QY	601 ALPGMWODRDVYLOGPIWAKIPHTDGHFSPPLMGFGGLKNPPEQILIKNTPVANPPA 660
Db	600 VLPGMWODRDVYLOGPIWAKIPHTDGHFSPPLMGFGGLKHPPEQILIKNTPVANPST 659
QY	661 EFSATKFASFITQYSTGVVEIEWELQKENSKRWNPEVOYTSNYAKSANVDFTVDNNG 720
Db	660 TFSAKFASFITQYSTGVVEIEWELQKENSKRWNPEIQTSMYNNKSVNVDFTVDITNGV 719
QY	721 YTEPRPIGTRYLTRPL 736
Db	720 YSEPRPIGTRYLTRNL 735
RESULT 8	
AAB50326	standard; protein; 734 AA.
AC	AAB50326;
DT	09-MAR-2001 (first entry)
XX	Adeno-associated virus capsid protein sequence.
XX	Adeno-associated virus; AAV; capsid; virus binding inhibition;
KM	competitive inhibitor.

XX Mastadenovirus.  
OS Synthetic.  
XX WO200073316-A2.  
XX 07-DEC-2000.  
PD 26-MAY-2000; 2000WO-US14466.  
PF 28-MAY-1999; 99US-0321589.  
XX  
PR (CELL-) CELL GENESYS INC.  
PA  
XX Patel S, McArthur J;  
PI WPI; 2001-061507/07.  
DR  
XX New polypeptide portion of a virus protein binding to an antibody  
PT specific for the virus useful for inhibiting binding of the virus to a  
PT cell or for binding host antibody to provide a transient tolerant or  
PT non-responsive state  
XX  
PS Disclosure; Fig 2; 33pp; English.  
XX  
CC The present sequence is given in a specification relating to  
CC polypeptide portions of a virus protein or its derivative, that bind to  
CC an antibody specific for the virus or inhibit binding of the virus to a  
CC cell. The polypeptides are used in inhibiting the binding of viruses to  
CC cells of a host. Oligopeptides that inhibit binding of virus to  
CC receptor can be used as competitive inhibitors to release bound virus  
CC in an adsorption-type assay, and if an antibody was used as an  
CC immunoadsorbent, the oligopeptide could be used to elute bound virus  
CC from a solid support to which virus antibody is immobilised. These  
CC oligopeptides may further be used to bind to host antibody to provide a  
CC transient tolerant or non-responsive state.  
XX  
SQ Sequence 734 AA;  
  
Query Match 84.6%; Score 3376; DB 22; Length 734;  
Best Local Similarity 83.0%; Pred. No. 1.7e-262;  
Matches 611; Conservative 51; Mismatches 72; Indels 2; Gaps 2;  
  
QY 1 MAADGYLPDWLEDNLSEGIREFWDLKPGAPKPKANQOKODDGRGLVLPGYKYLGPFGND 60  
Db 1 MAADGYLPDWLEDTLSEGIROQWTKLPGPPPKPAERHKDSDRGLVLPGYKYLGPFGND 60  
  
QY 61 KGEPVNAADAALAEHDKAYDQOLKAGDNPYLRYNHADAEEFERLQEDTSFGNLTGRAVFQ 120  
Db 61 KGEPVNEADAALAEHDKAYDRQLDSGDNPYLKYNHADAEEFERLKEEDTSFGNLTGRAVFQ 120  
  
QY 121 AKKRVLEPLGLVEEGAKTAPGKKRVEQSPQEPDSSSGIGTKGOQPAKKRLNFGQTGDS 180  
Db 121 AKKRVLEPLGLVEEPVKTAPGKKRVEHSPVEPDSSSGTGKAGQOPARKRLNFGQTGDAD 180  
  
QY 181 SVDPDQPLGEPPTPAVGPPTMASGGGAPMADNNEGADGVGNASGNWCHDSTWLGDRV 240  
Db 181 SVDPDQPLGQPPAAPSGLGTNTWATGSGAPMADNNEGADGVGNSSGNWCHDSTWMDRV 240  
  
QY 241 TTSTRTWALPTYNNHLYKQISSASTGASNDNHYFGYSTPWCYFEDENRFCHFSPRDWORL 300  
Db 241 TTSTRTWALPTYNNHLYKQISSQS-GASNDNHYFGYSTPWCYFEDENRFCHFSPRDWORL 299  
  
QY 301 INNNWGERPKRLNFKLFNIQVKEVTNTDGTITIANNLSTVOYFSDSEYQLPYVLGSAHQ 360  
Db 300 INNNWGERPKRLNFKLFNIQVKEVTNDGTITIANNLSTVOYETDSEYQLPYVLGSAHQ 359  
  
QY 361 GCLPPPADVEMIPQYGYTLNNGSGAVGRSSFYCLEYFPGQMLRTGNNFTFSYTFEEVP 420  
Db 360 GCLPPPADVEMIPQYGYTLNNGSGAVGRSSFYCLEYFPGQMLRTGNNFTFSYTFEDVP 419  
  
QY 421 FHSSYAHQSQSLDRMLNPLIDQYLYLNTQNGSGAQNKDLFSGSGSPAGMSVQPKMLP 480  
Db 421 FHSSYAHQSQSLDRMLNPLIDQYLYLNTQNGSGAQNKDLFSGSGSPAGMSVQPKMLP 480

Db 420 FHSSYAHQSQSLDRMLNPLIDQYLYLRTNTPSGTTTQSRLOFSQAGASDIRDQSRNWL 479  
QY 481 GPCYRQORVSKTKTDNNNSNFTWTGASKYNNLNGRESTINGTAMASHKDEDEKFFPMSCV 540  
Db 480 GPCYRQORVSKTSADNNNSSEYMTGATKYHLNGRDGLVNPBPAMASHKDEDEKFFPMSCV 539  
QY 541 MIFGKESAGASNTALDNVMTDEEIKATNPVATERFGTVAVNFQSSSTDTPATGVDYHMG 600  
Db 540 LIFGKQSEKTNVDIEKMTDDEEIRTN-VATEQYGSVSTNLQGRNQAATADVNTQG 598  
QY 601 ALPGMWQDRDVLQGPIMAKIPHTDGHFHPSPLMGFGIKNPPQILIKNTPVANPBA 660  
Db 599 VLPGMWQDRDVLQGPIMAKIPHTDGHFHPSPLMGFGIKNPPQILIKNTPVANPST 658  
QY 661 EFSATKFAFITQYSTGVVEIEMWLOKENSKRWNPEVQYTSNYAKSANVDFTVDNGL 720  
Db 659 TFSAKFAFITQYSTGVVEIEMWLOKENSKRWNPEIQTSTNYNKS VNVDFTVDNGLV 718  
QY 721 YTEPRPIGTRYLTRPL 736  
Db 719 YSEPRPIGTRYLTRNL 734  
  
RESULT 9  
AAY71168  
ID AAY71168 standard; Protein; 599 AA.  
XX  
AC AAY71168;  
XX  
DT 08-SEP-2000 (first entry)  
XX  
DE Adeno-associated virus serotype 1 capsid protein VP2.  
XX  
KW Adeno-associated virus serotype 1; AAV-1; rep protein; capsid protein;  
KW cap protein; recombinant viral vector; gene delivery; gene therapy;  
KW vaccine; transgene; VP2.  
XX  
OS Adeno associated virus serotype 1.  
XX  
PN WO200028061-A2.  
XX  
PD 18-MAY-2000.  
XX  
PF 02-NOV-1999; 99WO-US25694.  
XX  
PR 05-NOV-1998; 98US-0107114.  
XX  
PA (UYPE-) UNIV PENNSYLVANIA.  
XX  
PI Wilson JM, Xiao W;  
XX  
DR WPI; 2000-376571/32.  
DR N-PSDB; AAD00772, AAD00778.  
XX  
PT Novel adeno-associated virus serotype 1 polynucleotide useful for  
PT preparation of medicament for delivery of a transgene to a host  
XX  
PS Claim 7; Page 93-95; 108pp; English.  
XX  
CC The patent discloses an adeno-associated virus serotype 1 (AAV-1) DNA  
CC which is characterised by two inverted terminal repeats (ITR) and open  
CC reading frames for rep and capsid (cap) proteins. The rep reading frame  
CC encodes four proteins, Rep 78, Rep 68, Rep 52 and Rep 40, while the cap  
CC reading frame encodes three structural proteins, VP1, VP2 and VP3.  
CC The AAV-1 sequence or its fragments particularly ITRs, rep and cap  
CC coding regions, are useful in production of recombinant viral vectors  
CC for gene delivery. These vectors can be used as gene therapy  
CC vectors, vaccine vectors or antisense delivery vectors. The AAV-1 does  
CC not induce the formation of neutralising antibodies specific to any  
CC serotype of AAV hence is useful for transforming host cells, and in the  
CC preparation of a medicament for the delivery of transgene to a host.  
CC The present sequence is an AAV-1 cap protein VP2 which is  
CC useful in the production of recombinant viral vector for gene delivery.



XX	Sequence	599 AA;			
SQL	Query Match	81.5%;	Score 3251;	DB 21;	Length 599;
	Best Local Similarity	100.0%;	Pred. No. 1.4e-252;		
	Matches 599;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	138	TAPGKRPFVEQSPOEPDSSSGIGKTGOQPAKRLNFGQTGDSSESVDPDPLGEPATPAA	197		
DB	1	TAPGKRPFVEQSPOEPDSSSGIGKTGOQPAKRLNFGQTGDSSESVDPDPLGEPATPAA	60		
QY	198	VGPTMASGGAPMADNNEGADGVGNASGNWCHDSTWLGDRVITTTSTRTWALPTYNNHLY	257		
DB	61	VGPTMASGGAPMADNNEGADGVGNASGNWCHDSTWLGDRVITTTSTRTWALPTYNNHLY	120		
QY	258	KQISSASTGASNDNHYFGYSTPWGYPDFNRFCHFSPRDWQRLINNNGFRPKRLNFKLF	317		
DB	121	KQISSASTGASNDNHYFGYSTPWGYPDFNRFCHFSPRDWQRLINNNGFRPKRLNFKLF	180		
QY	318	NIQVKEVTNDGVTITIANNLSTVQVFSDEYQLPYVLGSAHQGCLPPFPADVFMIPQYG	377		
DB	181	NIQVKEVTNDGVTITIANNLSTVQVFSDEYQLPYVLGSAHQGCLPPFPADVFMIPQYG	240		
QY	378	YLTLLNGSQAVGRSSFYCLEYFPSSQMLRTGNNTFSYTFEEVPHSSYAHQSQSLDRLMP	437		
DB	241	YLTLLNGSQAVGRSSFYCLEYFPSSQMLRTGNNTFSYTFEEVPHSSYAHQSQSLDRLMP	300		
QY	438	LIDQYLYLNRTONQSGSAQNKDLFSRGSFAGMSVQPKNWLPGPCYRQORVSKTKTDNN	497		
DB	301	LIDQYLYLNRTONQSGSAQNKDLFSRGSFAGMSVQPKNWLPGPCYRQORVSKTKTDNN	360		
QY	498	NSNFTWTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMGVMIFGKESAGASNTALDN	557		
DB	361	NSNFTWTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMGVMIFGKESAGASNTALDN	420		
QY	558	VMITDEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMGALPGMWQDRDVLQGP	617		
DB	421	VMITDEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMGALPGMWQDRDVLQGP	480		
QY	618	IWAKIPHTDGHFHPSPLMGGFGLKNPPQILIKNTVPANPPAEFSATKFAFITQYSTG	677		
DB	481	IWAKIPHTDGHFHPSPLMGGFGLKNPPQILIKNTVPANPPAEFSATKFAFITQYSTG	540		
QY	678	QVSVEIEWELQKENSKRWNPEVQYTSNYAKSANVDFVDNNGLYTEPRPIGTRYLTRPL	736		
DB	541	QVSVEIEWELQKENSKRWNPEVQYTSNYAKSANVDFVDNNGLYTEPRPIGTRYLTRPL	599		
RESULT 10					
AAV71169	standard; Protein; 534 AA.				
ID	AAV71169 standard; Protein; 534 AA.				
XX					
AC	AAV71169;				
XX					
DT	08-SEP-2000 (first entry)				
XX					
DE	Adeno-associated virus serotype 1 capsid protein VP3.				
XX					
KW	Adeno-associated virus serotype 1; AAV-1; rep protein; capsid protein;				
KW	cap protein; recombinant viral vector; gene delivery; gene therapy;				
KW	vaccine; transgene; VP3.				
XX					
OS	Adeno associated 'virus serotype 1.				
XX					
PN	MO200028061-A2.				
XX					
PD	18-MAY-2000.				
XX					
PF	02-NOV-1999; 99WO-US25694.				
XX					
PR	05-NOV-1998; 98US-0107114.				
XX					
PA	(UYPE-) UNIV PENNSYLVANIA.				

XX	Wilson JM, Xiao W;				
XX	WIPI; 2000-376571/32.				
DR	N-PSDB; AAD00772, AAD00779.				
XX					
PT	Novel adeno-associated virus serotype 1 polynucleotide useful for				
PT	preparation of medicament for delivery of a transgene to a host				
XX	Claim 7; Page 99-101; 108pp; English.				
XX					
CC	The patent discloses an adeno-associated virus serotype 1 (AAV-1) DNA				
CC	which is characterised by two inverted terminal repeats (ITR) and open				
CC	reading frames for rep and capsid (cap) proteins. The rep reading frame				
CC	encodes four proteins, Rep 78, Rep 68, Rep 52 and Rep 40, while the cap				
CC	reading frame encodes three structural proteins, VP1, VP2 and VP3.				
CC	The AAV-1 sequence or its fragments particularly ITRs, rep and cap.				
CC	coding regions, are useful in production of recombinant viral vectors				
CC	for gene delivery. These vectors can be used as gene therapy				
CC	vectors, vaccine vectors or antisense delivery vectors. The AAV-1 does				
CC	not induce the formation of neutralising antibodies specific to any				
CC	serotype of AAV hence is useful for transforming host cells, and in the				
CC	preparation of a medicament for the delivery of transgene to a host.				
CC	The present sequence is an AAV-1 cap protein VP3 which is				
CC	useful in the production of recombinant viral vector for gene delivery.				
SQL	Sequence	534 AA;			
Query Match	72.9%;	Score 2906;	DB 21;	Length 534;	
Best Local Similarity	100.0%;	Pred. No. 6.8e-225;			
Matches 534;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
QY	203	MASGGAPMADNNEGADGVGNASGNWCHDSTWLGDRVITTTSTRTWALPTYNNHLYKQISS	262		
DB	1	MASGGAPMADNNEGADGVGNASGNWCHDSTWLGDRVITTTSTRTWALPTYNNHLYKQISS	60		
QY	263	ASTGASNDNHYFGYSTPWGYPDFNRFCHFSPRDWQRLINNNGFRPKRLNFKLFNIQYK	322		
DB	61	ASTGASNDNHYFGYSTPWGYPDFNRFCHFSPRDWQRLINNNGFRPKRLNFKLFNIQYK	120		
QY	323	EVTNDGVTITIANNLSTVQVFSDEYQLPYVLGSAHQGCLPPFPADVFMIPQYGYLTIN	382		
DB	121	EVTNDGVTITIANNLSTVQVFSDEYQLPYVLGSAHQGCLPPFPADVFMIPQYGYLTIN	180		
QY	383	NGSQAVGRSSFYCLEYFPSSQMLRTGNNTFSYTFEEVPHSSYAHQSQSLDRLMPIDQY	442		
DB	181	NGSQAVGRSSFYCLEYFPSSQMLRTGNNTFSYTFEEVPHSSYAHQSQSLDRLMPIDQY	240		
QY	443	LYYLNRTONQSGSAQNKDLFSRGSFAGMSVQPKNWLPGPCYRQORVSKTKTDNNNSNFT	502		
DB	241	LYYLNRTONQSGSAQNKDLFSRGSFAGMSVQPKNWLPGPCYRQORVSKTKTDNNNSNFT	300		
QY	503	WTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMGVMIFGKESAGASNTALDNVMTTD	562		
DB	301	WTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMGVMIFGKESAGASNTALDNVMTTD	360		
QY	563	EEBIIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMGALPGMWQDRDVLQGPWAKI	622		
DB	361	EEBIIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMGALPGMWQDRDVLQGPWAKI	420		
QY	623	PHTDGHFHPSPLMGGFGLKNPPQILIKNTVPANPPAEFSATKFAFITQYSTGQVSVE	682		
DB	421	PHTDGHFHPSPLMGGFGLKNPPQILIKNTVPANPPAEFSATKFAFITQYSTGQVSVE	480		
QY	683	IEMELQKENSKRWNPEVQYTSNYAKSANVDFVDNNGLYTEPRPIGTRYLTRPL	736		
DB	481	IEMELQKENSKRWNPEVQYTSNYAKSANVDFVDNNGLYTEPRPIGTRYLTRPL	534		
RESULT 11					
AA65793	standard; Protein; 598 AA.				
ID	AA65793 standard; Protein; 598 AA.				
XX					

AC	AAG65793;	
XX		
DT	11-FEB-2002	(first entry)
XX		
DE	Adeno-associated virus 2 (AAV-2)	major coat protein VP2.
XX		
KW	Recombinant viral vector; RVV;	capsid; parvovirus; transgene; cytostatic;
KW	inverted terminal repeat; nootropic;	neuroprotective; antianemic; ITR;
KW	antidiabetic; antitumour; gene	therapy; adeno-associated virus; AAV;
KW	major coat protein; AAV-2;	VP2.
XX		
OS	Adeno-associated virus 2.	
XX		
FH	Key	Location/Qualifiers
FT	Misc-difference 1	
FT	/note= "encoded by ACG"	
XX		
PN	WO200168888-A2.	
XX		
PD	20-SEP-2001.	
XX		
PF	13-MAR-2001;	2001WO-US07927.
XX		
PR	14-MAR-2000;	2000US-189110P.
XX		
PA	(NEUR-) NEUROLOGIX INC.	
XX		
PI	Xiao W, During MJ;	
XX		
DR	WPI; 2001-536912/67.	
DR	N-PSDB; AAI66974.	
XX		
PT	Recombinant viral vector useful	in improving gene therapy in a subject,
PT	and for increasing efficiency of	entry into a cell, comprises a
PT	chimeric capsid having one non-	native amino acid sequence and a
PT	desired transgene	
XX		
PS	Disclosure; Page, 51; 53pp;	English.
XX		
CC	The invention provides a recombinant viral vector (RVV) comprising a	
CC	chimeric capsid (I) having at least one non-native amino acid sequence,	
CC	derived from a capsid protein domain of parvovirus (II), a virus (III),	
CC	or their combination, and a transgene flanked 5' and 3' by inverted	
CC	terminal repeat (ITR) sequences, derived from (II), (III), or their	
CC	combination. The RVV is useful for improving gene therapy in a subject	
CC	with a disorder, and for increasing the efficiency of entry into a cell,	
CC	which involves producing (I) encapsulating a viral vector, and contacting	
CC	a cell with RVV having (I) such that (I) binds to an attachment site on	
CC	the cell surface and permits the vector to enter the cell efficiently.	
CC	A pharmaceutical composition comprising RVV with (I) containing a	
CC	transgene sequence associated with a disease or a disorder such that	
CC	expression of the transgene would result in amelioration of the disease	
CC	or disorder such as inherited neurological and metabolic diseases e.g.	
CC	lysosomal storage disease, Lesch-Nyhan syndrome, amyloid polynuropathy,	
CC	Alzheimer amyloid, Duchenne's muscular dystrophy, diseases of the blood	
CC	e.g. sickle-cell anemia, clotting disorders and thalassemias, cystic	
CC	fibrosis, diabetes, diseases associated with hormone deficiencies,	
CC	retinoblastoma and various types of neoplastic cells which include	
CC	tumours especially central nervous system tumours, neoplasms, carcinomas,	
CC	sarcomas, leukemias and lymphoma. The present sequence represents the	
CC	adeno-associated virus 2 (AAV-2) major coat protein VP2. AAV-2 sequences	
CC	are used in the construction of a chimeric vector.	
XX		
Sequence	598	AA;
XX		

Query Match	69.2%;	Score 2759.5;	DB 22;	Length 598;
Best Local Similarity	82.8%;	Pred. No. 4.9e-213;		
Matches 495;	Conservative 43;	Mismatches 59;	Indels 1;	Gaps 1;

```
Qy      139 APEGKRVEQSPOEPDSSSGIGKTGOQPAKKRLNFQTGDSSEVPDPQLGEBPATPAAV 198
        ||||| | | | | | : | | | | | : | | | | | : | | :
Db       2  APEGKRPVESHSPVEDSSSGTGAKGOQPAKKRLNFGQTGDADSVDPQLGQPAPASGL 61
```

Qy	199	GPTTMA	SCGCAPMADNNEGADGVGNASGNWHCDSTWLGDRVITTTSTRWALPTYNNHLYK	258
Dd	62	GTNTMATGSGAPMADNNEGADVGNSSGNWHCDSTMGDRIVITTTSTRWALPTYNNHLYK		121
Qy	259	QISSASTGASNDNHYFGYSTPMGYEDFNREHFCHFSPRDWORLLINNNGFBRKLNFLEN		318
Dd	122	QISSQS-GASNDNHYFGYSTPMGYEDFNRFHCHFSPRDWORLLINNNGFBRKLNFLEN		180
Qy	319	IQYKEVTINDGVTTIANNLTSIVQYFSDSEYQLPYVLGSAHQCLPPPADVFMIPOYGX		378
Dd	181	IQYKEVTQNDGTTTIANNLTISTVOYFTDSEYQLPYVLGSAHQCLPPPADVFEMVPOYGX		240
Qy	379	LTLNNGSQAVGRSSFYCLEYFPSPOMLRTGNNFTFSYTFEEVPFHSSYASHOSLDRLMNPL		438
Dd	241	LTLNNGSQAVGRSSFYCLEYFPSPOMLRTGNNFTFSYTFEDEVPFHSSYASHOSLDRLMNPL		300
Qy	439	IDOYLYYLNRTONQSGSAQNKDLESRGSPAGMSVOPKNWLPGPCYRQORVSKTKTDNNN		498
Dd	301	IDOYLYYLSRTNTPSGTITTOQRLOFSQA GASDIRDO SRNWLPGPCYRQORVSKTSADNNN		360
Qy	499	SNFTWTGASKYNLNGRESINPGTAMASHKHDEDEKFFPMSGVMIFGESAGASANTALDNV		558
Dd	361	SEYSWTGATKYHLNGRDSLNVNPGPAMASHKHDEEKFFPQSGVLIIFGQOSEKTNV DIEKV		420
Qy	559	MITDEEEKATNPVATERFGTVAVNFOSSSTD PATGDVHAMGALPGMWODRDVYLQGPI		618
Dd	421	MITDEEIRRTNPNVATEQYGSVS TNLQGRNRQAATA DAVNTQGVLPENWQDRDVYLQGPI		480
Qy	619	WAKI PHTDGHFHPSPLMGGFGLKNP PPQILIKNTPV PANP PAEF SATKFA SFITQYSTGO		678
Dd	481	WAKIPHTDGHFHPSPLMGFGFLKHP PPQILIKNTPV PANP STFSAKFA SFITQYSTGO		540
Qy	679	VSVEIEWELOKENS KRWNPEVOYT SNYAKS ANVDFTVDNNGLYTEBPPIGTRYLTRPL		736
Dd	541	VSVEIEWELOKENS KRWNPELO YTSNYNKS VNVDFTVDINGVYSEBPPIGTRYLTRNL		598

CC	The present sequence is provided in a specification relating to a virus-like particle-forming protein, and to a peptide containing a nuclear-shifting signal at its N-terminus that can form virus-like
PS	Claim 1; Page 14-16; 33pp; Japanese.
XX	
XX	New virus-like particles from VP3 capsid protein of adeno-associated
PT	virus, comprise a peptide containing a nucleus-shifting signal
PT	connected to its N-terminal -
XX	
PA	(HAND/) HANDA H.
XX	
DR	WPI; 2001-599854/68.
XX	
XX	26-JUN-2001.
XX	
PF	30-JUL-1999; 99JP-0249140.
XX	
PR	30-JUL-1999; 99JP-0249140.
XX	
PA	(HAND/) HANDA H.
XX	
OS	Adeno associated virus.
XX	
PN	JP2001169777-A.
XX	
PD	26-JUN-2001.
XX	
XX	30-JUL-1999; 99JP-0249140.
XX	
XX	30-JUL-1999; 99JP-0249140.
XX	
DE	Adeno-associated virus VP2 capsid protein.
XX	
XX	Adeno-associated virus; AAV; VP2; capsid; virus-like particle;
KM	nuclear localisation signal; VP3.
XX	
DT	02-JAN-2002 (first entry)
AC	AAM51509;
XX	
ID	AAM51509 standard; protein; 598 AA.
XX	
XX	AAM51509
XX	
RESULT 12	

CC particles by shifting to the nucleus of the animal cell in which it  
CC is expressed. The method is used for forming virus-like particles  
CC from the VP3 capsid protein of adeno-associated virus (AAV).

XX Sequence 598 AA;

Query Match 69.2%; Score 2759.5; DB 22; Length 598;  
Best Local Similarity 82.8%; Pred. No. 4.9e-213;  
Matches 495; Conservative 43; Mismatches 59; Indels 1; Gaps 1;

QY 139 APGKRPEVQSPQEPDSSSGIGKTGQOPAKKRLNFGQTGDSSEVPDPQLGEPATPAV 198  
DB 2 APGKRPEVHSPEVDPDSSSGTGKAGQOPARKRLNFGQTGDADSVDPDQLGQPPAAPSG 61  
QY 199 GPTTMASGGAPMADNNEGADGVGNASGNWCHDSTWLGDRVITTSRTWALPTYNNHLYK 258  
DB 62 GTNTMATGSGAPMADNNEGADGVGNSSGNWCHDSTWMDRVITTSRTWALPTYNNHLYK 121  
QY 259 QISSASTGASNDNHFGYSTPWGYFDENRFHCHFSPRDWQRLINNNGFRPKRLNFKLFN 318  
DB 122 QISSQS-GASNDNHFGYSTPWGYFDENRFHCHFSPRDWQRLINNNGFRPKRLNFKLFN 180  
QY 319 IQVKEVTNDGVTTIANNLTSTVQVFSDEYQLPYVLGSAHQGLPPPADVEMIPQYGY 378  
DB 181 IQVKEVTNDGVTTIANNLTSTVQVFTDSEYQLPYVLGSAHQGLPPPADVEMIPQYGY 240  
QY 379 LTLNNGSQAVGRSSFCLEYFSPQMLRTGNNFTFSYTFEEVEFHSSVAHSQSLDRLMPL 438  
DB 241 LTLNNGSQAVGRSSFCLEYFSPQMLRTGNNFTFSYTFEDVFPFSSVAHSQSLDRLMPL 300  
QY 439 IDQYLYLNRTQNSQSAQNKDLFSRGSFAGMSVQPKNWLPGPCYRQQRVSKTKTDNN 498  
DB 301 IDQYLYLSRTNPSGTTTQSRLOFSQAGASDIRDQSRNWLPGPCYRQQRVSKTSADNN 360  
QY 499 SNFTWTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMGVMIFGKESAGASNTALDNV 558  
DB 361 SEYSMTGATKYHLNGRDSLVPNGPAMASHKDEDEKFFPQSGVLIFGKQSEKTNVDIEKV 420  
QY 559 MITDEEIKATNPVATERFGTVAVNFQSSSTDPATGVDHAMGALPGMWQDRDVLQGP 618  
DB 421 MITDEEIRTTNPVATEQYGSVSTNLQGRNQAATADVNTQGVLPGMWQDRDVLQGP 480  
QY 619 WAKIPHTDGHFSPPLMGFGLKNPPQILLIKNTVPANPPAFAFSATKFAFITQYSTGQ 678  
DB 481 WAKIPHTDGHFSPPLMGFGLKHPPIQLIKNTVPANPSTTFSAAKFAFITQYSTGQ 540  
QY 679 VSVEIEMELQKENSRRNPEVOYTSNYAKSANDFTVDNNGLYTEPRPIGTRYLTRPL 736  
DB 541 VSVEIEMELQKENSRRNPEIQTSTNYNKSANDFTVDNGVYSEPRPIGTRYLTRNL 598

RESULT 13

AAW46308 ID AAW46308 standard; Protein; 734 AA.

XX AC AAW46308;

XX DT 28-AUG-1998 (first entry)

XX DE AAV4 VP1 capsid protein.

XX KM AAV; AAV-4; VP1; capsid protein; vector; gene transfer;

XX KW gene delivery; cancer; gene therapy.

XX OS Adeno associated virus 4.

XX FH Key Location/Qualifiers

FT Misc-difference 208 /note= "encoded by GGC"

FT Misc-difference 250 /note= "encoded by CTN"

FT /note= "encoded by CTN"

XX PN WO9811244-A2.

XX 19-MAR-1998.  
XX 11-SEP-1997; 97WO-US16266.  
XX 11-SEP-1996; 96US-0025934.  
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
XX Chlorini JA, Kotin RM, Safer B;  
XX MPI; 1998-207403/18.  
XX N-PSDB; AAV21648 AND AAV21650.

PT Adeno-associated virus 4 based vectors - used to transduce erythroid  
PT progenitor cells for treatment of e.g. cancer and other genetic  
PT diseases

PS Claim 34; Page 54-55; 80pp; English.

CC This polypeptide comprises the adeno associated virus 4 (AAV4)  
CC VP1 capsid protein that is encoded by open reading frame 2 (see  
CC AAV21650) of the AAV4 genome (see also AAV21648). The invention also  
CC provides AAV4 VP2 (see AAW46313) and VP3 (see AAW46314), also encoded  
CC by the AAV4 genome (see also AAV21657-58). The invention provides  
CC recombinant vectors and viral particles based on AAV4 that may be  
CC useful for transducing erythroid progenitor cells for the treatment  
CC of e.g. cancer and genetic diseases which can be corrected by bone  
CC marrow transplants using matched donors. A claimed method of  
CC delivering a nucleic acid to a subject (including a subject with  
CC antibodies to AAV2) comprises administering to a cell from the  
CC subject an AAV4 particle comprising the nucleic acid inserted  
CC between a pair of AAV inverted terminal repeats (see AAV21651 and  
CC AAV21659), and returning the cell to the subject.

XX Sequence 734 AA;

Query Match 62.3%; Score 2486.5; DB 19; Length 734;  
Best Local Similarity 63.5%; Pred. No. 6e-191;  
Matches 475; Conservative 81; Mismatches 161; Indels 31; Gaps 10;

QY 4 DGYLPWLEDNLSEGIREWMDLKPAPKPKANOQKQDDGRGLVLPYKYLGPFNGLDKGE 63  
DB 3 DGYLPWLEDNLSEGVREWMALQPGAPKPKANOQHQNARGLVLPYKYLGPGNGLDKGE 62  
QY 64 PVNAADAAALEHDKAYDQQLKAGDNPYLRYNHADAEPQERLOEDTSFGNLRGAVFOAKK 123  
DB 63 PVNAADAAALEHDKAYDQQLKAGDNPYLRYNHADAEPQERLOEDTSFGNLRGAVFOAKK 122  
QY 124 RVLPEPLGVEEAKTAPGKRPVEQSPQEPDSSSGIGKTGQOPAKKRLNF-GQTGDSSEV 182  
DB 123 RVLPEPLGVEQAGETAPGKRPVIESPQPDSSSTGIGKKGQPAKKLVFEDETGAGDGP 182  
QY 183 PDPQLGEPATPAVGPPTMASGGA.PMADNNEGADGVGNASGNWCHDSTWLGDRVITTT 242  
DB 183 PEGSTSG-----AMSDSEMRRAAGAAGAVEGGQADGVNASGDWCHDSTWSEGHVTTT 236  
QY 243 STRTWALPTYNNHLYKQISSASTGASNDNHFGYSTPWGYFDENRFHCHFSPRDWQRLIN 302  
DB 237 STRTWALPTYNNHLYKRLGE----SLQSNNTYNGFSTPWGYFDENRFHCHFSPRDWQRLIN 292  
QY 303 NNWGFPRKRLNFKLFNIQVKEVTNDGVTTIANNLTSTVQVFSDEYQLPYVLGSAHQGC 362  
DB 293 NNWGMRPKAMRVKIFNIQVKEVTTSGETTVAANNLTSTVQVIFADSSYELPYMDAGQEGS 352  
QY 363 LPPPADVEMIPQYGY---LTLNNGSQAVGRSSFCLEYFSPQMLRTGNNFTFSYTFEEV 419  
DB 353 LPPPADVEMIPQYGYCGLVGTNTSQQOTDRNAFYCLEYFSPQMLRTGNNFEITYSFEKV 412  
QY 420 PFHSSVAHSQSLDRLMPLIDQYLYLNRTON---QSGSAQNKDLFSRGSFAGMSVOP 475  
DB 413 PFHSMVAHSQSLDRLMPLIDQYLYLWGLQSTTTGTTLNAGTATN---FTKLRTPTNFSNFX 469



[illegible]

RESULT 14  
 ID AAG65794 standard; Protein; 533 AA.  
 AC AAG65794;  
 DT 11-FEB-2002 (first entry)  
 DE Adeno-associated virus 2 (AAV-2) major coat protein VP3.  
 KW Recombinant viral vector; RVV; capsid; parvovirus; transgene; cytosstatic;  
 KW inverted terminal repeat; nootropic; neuroprotective; antianemic; ITR;  
 KW antidiabetic; antitumour; gene therapy; adeno-associated virus; AAV;  
 KW major coat protein; AAV-2; VP3.  
 OS Adeno-associated virus 2.  
 PN WO200168888-A2.  
 PD 20-SEP-2001.  
 PF 13-MAR-2001; 2001WO-US07927.  
 PR 14-MAR-2000; 2000US-189110P.  
 PA (NEUR-) NEUROLOGIX INC.  
 PI Xiao W, During MJ;  
 DR WPI; 2001-596912/67.  
 DR N-PSDB; AAI66974.  
 PT Recombinant viral vector useful in improving gene therapy in a subject,  
 PT and for increasing efficiency of entry into a cell, comprises a  
 PT chimeric capsid having one non-native amino acid sequence and a desired  
 PT transgene -  
 PS Disclosure; Page 51; 53pp; English.  
 CC The invention provides a recombinant viral vector (RVV) comprising a  
 CC chimeric capsid (I) having at least one non-native amino acid sequence,  
 CC derived from a capsid protein domain of parvovirus (II), a virus (III),  
 CC or their combination, and a transgene flanked 5' and 3' by inverted  
 CC terminal repeat (ITR) sequences, derived from (II), (III), or their  
 CC combination. The RVV is useful for improving gene therapy in a subject  
 CC with a disorder, and for increasing the efficiency of entry into a cell,  
 CC which involves producing (I) encapsulating a viral vector, and contacting  
 CC a cell with RVV having (I) such that (I) binds to an attachment site on  
 CC the cell surface and permits the vector to enter the cell efficiently.  
 CC A pharmaceutical composition comprising RVV with (I) containing a  
 CC transgene sequence associated with a disease or a disorder such that  
 CC expression of the transgene would result in amelioration of the disease

CC or disorder such as inherited neurological and metabolic diseases e.g.  
CC lysosomal storage disease, Leisch-Nyhan syndrome, amyloid polynuropathy,  
CC Alzheimer amyloid, Duchenne's muscular dystrophy, diseases of the blood  
CC e.g. sickle-cell anemia, clotting disorders and thalassemias, cystic  
CC fibrosis, diabetes, diseases associated with hormone deficiencies,  
CC retinoblastoma and various types of neoplastic cells which include  
CC tumours especially central nervous system tumours, neoplasms, carcinomas,  
CC sarcomas, leukemias and lymphoma. The present sequence represents the  
CC adeno-associated virus 2 (AAV-2) major coat protein VP3. AAV-2 sequences  
CC are used in the construction of a chimeric vector.

**SQ Sequence 533 AA;**

Query Match	62.2%	Score 2481.5;	DB 22;	length 533;
Best Local Similarity	83.3%	Pred. No. 9.4e-191;		
Matches 445; Conservative	37;	Mismatches 51;	Indels 1;	Gaps 1;

```

Oy      203 MASGGGAPMADNNEGADGVGNASGNHCHDSTWLGDRVITTTSTRTWALPTYNNHLYKQISS 2622
      ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db      1 MATGSGAPMADNNEGADGVGNSSGNNHCHDSTWMDGRVITTTSTRTWALPTYNNHLYKQISS 60

```

[illegible]

Qy	323	EVTINDGVTTIAN	LTSTQVFSDE	SYQLPYL	GSAGQGLPP	FPADVEMIPQY	GYLTIN	3822
Db	120	EVTQNDGTTIAN	LTSTQVFSDE	SYQLPYL	GSAGQGLPP	FPADVEMIPQY	GYLTIN	1799

Qy	383	NGSQAVGRSSFYCLEYFP	SQLRTGNFTFSYTFEEV	PFHSSYAH	SQSLDR	LNPLID	QY	442
Db	180	NGSQAVGRSSFYCLEYFP <th>SQLRTGNFTFSYTFEEV</th> <th>PFHSSYAH</th> <th>SQSLDR</th> <th>LNPLID</th> <td>QY</td> <td>239</td>	SQLRTGNFTFSYTFEEV	PFHSSYAH	SQSLDR	LNPLID	QY	239

```
Oy      443 LYYLNRTONOGSAQNKDLFSGSPAGMSVQPKNWLPGPCYRQORVSKTKTDNNNSFT 5022
        |||::|.||:||||:|:|||||:|||||||:|||||:|:
Db      240 LYYLSRTNTPSGTTTOSRLQFSQAGASDIRDQSRRNWLPGPCYRQORVSKTSADNNNSEYS 2998
```

```
Qy      503 WTGASKYNLNGRESIIINPGTAMASHKODEDEKFFPMGVMIFGKESAGASANTALDNWITD 5622
      ||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db      300 WTGATKYHLNGRDSLVPGPAMASHKODEEKFPPQSGVLI FGKQSEKTNVDIEKWMITD 3599
```

Qy	Db
563	360
EEBIKATNPVATERFEGTVAVNFQSSSTDPEATGDVHAMGALPGMWQDRDVIYLOGPWAKI	EEIIRTNPVATEQYGSVSTNLQGRNQATADVNTQGVLPGMWQDRDVIYLOGPWAKI
6222	4199

```
Qy      623 PHTDGHFHPSPLMGGFGLKNPPQILIKNTVPANPPAEFSATKFASTIQYSTGQVSE 682
        |||||:|||||
Db      420 PHTDGHFHPSPLMGGFGLKNPPQILIKNTVPANPSTTFSAAKFASTIQYSTGQVSE 479
```

```
Qy      683 IEMELOENSKRWNPENVQYTSNYAKSANVDFTVDNGGLYTEPRPIGTRYLTRLPL 736  
        |||||  
        :|||  
Db      480 IEMELOENSKRWNPFIQYTSNMYKSVNVDFTVDTNGVGYSEPRPIGTRYLTRLNL 533
```

RESULT 15  
AAM51510  
ID AAM51510 standard; protein; 533 AA

AC AAM51510;

DT 02-JAN-2002 (first entry)

DE Adeno-associated virus VP3 capsid protein.

KW Adeno-associated virus; AAV; capsid; virus-like particle;

OS Adeno associated virus.

PN JP2001169777-A.

PD 26-JUN-2001.

XX

```

Oy      203 MASGGGAPMADNNEGADGVGNASGNHCHDSTWLGDRVITTTSTRTWALPTYNNHLYKOISS 2622
      ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db      1  MATGSGAPMADNNEGADGVGNSSGSHCHDSTWMDRVITTTSTRTWALPTYNNHLYKOISS 60

```

Qy	263	ASTGASNDNHYFGYSTPMWGYFDENRHHCHFSPRDWÖRLINNMGFRPKRLNFKLENIÖVK   	3222
Db	61	QS-GASNDNHYFGYSTPMWGYFDENRHHCHFSPRDWÖRLINNMGFRPKRLNFKLENIÖVK   	1199

Qy	323	EVTINDGVTTIAN	LTSTQVFSDE	YQLPYVLS	AHQGCLPP	FPADVEMIPQ	GYLTIN	3822
			:			:		
Db	120	EVTQNDGTTIAN	LTSTQVFTDSE	YQLPYVLS	AHQGCLPP	FPADVEMIPQ	GYLTIN	1799

Oy		383	NGSQAVGRSSFYCLEYPFSQMLRTGNPFSTYTEEVPHSSYAHQSOLDRLMNP.LIDöY	4422
			:	
Db		180	NGSQAVGRSSFYCLEYPFSQMLRTGNPFSTYTEEDVPFHSSYAHQSOLDRLMNP.LIDöY	2399

```
Oy      443  LYYLNRTONOGSAQNKDLLFSRGSBAGMSVOPKNWLPGPCYRQORVSKTKTDNNNSNFT 502
         |||::|||.:::||:||||:|||||||:|||||||:|||||||:|||||||:|||||||:
Db      240  LYYLSRTNTPSGTTTOSRLQFSQAGASDIRDOSRNWLPGPCYRQORVSKTSADNNNSEYS 299
```

```
Qy      503 WTGASKYNLNGRESIIINPGTAMASHKODEDEKFFPMGVMIFGKESAGASANTALDNWITD 5622
      ||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db      300 WTGATKYHLNGRDSLVPGPAMASHKODEEKFPPQSGVLI FGKQSEKTNVDIEKWMITD 3599
```

Qy	Db
563	360
EEBIKATNPVATERFEGTVAVNFQSSSTDPEATGDVHAMGALPGMWQDRDVIYLOGPWAKI	EEIIRTNPVATEQYGSVSTNLQGRNQATADVNTQGVLPGMWQDRDVIYLOGPWAKI
6222	4199

```
Qy      623 PHTDGHFHPSPLMGGFGLKNPPQILIKNTVPANPPAEFSATKFASTIQYSTGQVSE 682
        |||||:|||||
Db      420 PHTDGHFHPSPLMGGFGLKNPPQILIKNTVPANPSTTFSAAKFASTIQYSTGQVSE 479
```

```
Qy      683 IEMELOENSKRWNPENVQYTSNYAKSANVDFTVDNGGLYTEPRPIGTRYLTRLPL 736  
        |||||  
        :|||  
Db      480 IEMELOENSKRWNPFIQYTSNMYKSVNVDFTVDTNGVGYSEPPRIGTRYLTRLNL 533
```

RESULT 15  
AAM51510  
ID AAM51510 standard; protein; 533 AA

AC AAM51510;

DT 02-JAN-2002 (first entry)

DE Adeno-associated virus VP3 capsid protein.

KW Adeno-associated virus; AAV; capsid; virus-like particle;

OS Adeno associated virus.

PN JP2001169777-A.

PD 26-JUN-2001.

XX



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 2003, 18:29:05 ; Search time 18.1145 Seconds  
(without alignments)  
1195.466 Million cell updates/sec

Title: US-09-807-802A-13  
Perfect score: 3989  
Sequence: 1 MAADGYLPDWLEDNLSEGIR.....NNGLYTEPRPIGTRYLTRPL 736

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/iaa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2486.5	62.3	734	US-09-532-594B-4	Sequence 4, Appli
2	1830.5	45.9	598	US-09-532-594B-16	Sequence 16, Appl
3	1690.5	42.4	544	US-09-532-594B-18	Sequence 18, Appl
4	479.5	12.0	543	US-08-856-841-22	Sequence 22, Appl
5	430.5	10.8	500	US-08-856-841-16	Sequence 16, Appl
6	430.5	10.8	501	US-08-856-841-18	Sequence 18, Appl
7	429.5	10.8	486	US-08-856-841-19	Sequence 19, Appl
8	326	8.2	415	US-08-856-841-20	Sequence 20, Appl
9	318	8.0	395	US-08-856-841-13	Sequence 13, Appl
10	310	7.8	264	US-08-856-841-14	Sequence 14, Appl
11	303	7.6	398	US-08-856-841-21	Sequence 21, Appl
12	284	7.1	387	US-08-856-841-17	Sequence 17, Appl
13	216.5	5.4	579	US-09-532-594B-4	Sequence 4, Appli
14	192	4.8	584	US-09-532-594B-16	Sequence 16, Appl
15	142	3.6	210	US-08-856-841-9	Sequence 9, Appli
16	142	3.6	227	US-08-856-841-15	Sequence 15, Appl
17	142	3.6	250	US-08-856-841-12	Sequence 12, Appl
18	124	3.1	3060	US-08-487-826B-14	Sequence 14, Appl
19	119.5	3.0	1394	US-08-296-791-2	Sequence 2, Appli
20	119.5	3.0	1394	PCT-US95-10661A-2	Sequence 2, Appli
21	114.5	2.9	434	US-08-710-249-4	Sequence 4, Appli
22	114.5	2.9	434	US-09-220-157A-4	Sequence 2, Appli
23	113.5	2.8	655	US-08-469-202-27	Sequence 27, Appl
24	113.5	2.8	655	US-08-484-434C-34	Sequence 34, Appl
25	112.5	2.8	1096	US-09-415-946-3	Sequence 3, Appli
26	110.5	2.8	624	US-08-947-965-78	Sequence 78, Appl
27	110.5	2.8	655	US-08-469-202-28	Sequence 28, Appl

28	110.5	2.8	655	2	US-08-484-434C-35	Sequence 35, Appl
29	110.5	2.8	1751	4	US-09-136-574A-44	Sequence 44, Appl
30	109.5	2.7	1651	4	US-09-540-245A-18	Sequence 18, Appl
31	109.5	2.7	1848	4	US-08-296-791-6	Sequence 6, Appli
32	109.5	2.7	1848	5	PCT-US95-10661A-6	Sequence 6, Appli
33	109	2.7	1186	1	US-08-485-568A-4	Sequence 4, Appli
34	109	2.7	1186	1	US-08-357-698-6	Sequence 6, Appli
35	109	2.7	1186	2	US-08-590-554A-4	Sequence 4, Appli
36	109	2.7	1186	2	US-09-184-223-4	Sequence 4, Appli
37	109	2.7	1186	5	PCT-US93-12682-6	Sequence 6, Appli
38	109	2.7	1545	4	US-08-296-791-4	Sequence 4, Appli
39	109	2.7	1545	5	PCT-US95-10661A-4	Sequence 4, Appli
40	107.5	2.7	331	3	US-08-793-426A-2	Sequence 2, Appli
41	107.5	2.7	331	4	US-09-294-565-2	Sequence 2, Appli
42	106.5	2.7	331	1	US-08-136-993-1	Sequence 1, Appli
43	106.5	2.7	331	3	US-09-109-063-1	Sequence 1, Appli
44	106.5	2.7	331	3	US-08-793-426A-3	Sequence 3, Appli
45	106.5	2.7	331	4	US-09-294-565-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1  
US-09-532-594B-4  
; Sequence 4, Application US/09532594B  
; Patent No. 6468524  
; GENERAL INFORMATION:  
; APPLICANT: Chorini, John A.  
; APPLICANT: Kotin, Robert M.  
; APPLICANT: Safir, Brian  
; APPLICANT: Davidson, Beverly  
; TITLE OF INVENTION: AAV4 VECTOR AND USES THEREOF  
; FILE REFERENCE: 14014.0252U2  
; CURRENT APPLICATION NUMBER: US/09/532,594B  
; CURRENT FILING DATE: 2000-03-22  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 734  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence; No. 6468524e =  
; NAME/KEY: misc feature  
; OTHER INFORMATION: AAV4 capsid protein VP1  
; US-09-532-594B-4

Query Match 62.3%; Score 2486.5; DB 4; Length 734;  
Best Local Similarity 63.5%; Pred. No. 2.5e-209;  
Matches 475; Conservative 81; Mismatches 161; Indels 31; Gaps 10;

QY	4	DGYLPDWLEDNLSEGI	REWWDLKPGAPKPKANQKQDDGRGLVLP	PGYKYLGPFGDLKGE	63
Db	3	DGYLPDWLEDNLSEGI	REWWDLKPGAPKPKANQKQDDGRGLVLP	PGYKYLGPFGDLKGE	62
QY	64	PVNAADAAAL	EHDKAYDQOLKAGDNPYLRYNHADA	FEORLOEDTSFGN	LGRAVQAKK 123
Db	63	PVNAADAAAL	EHDKAYDQOLKAGDNPYLRYNHADA	FEORLOEDTSFGN	LGRAVQAKK 122
QY	124	RVLEPLGLVEE	GAKTAPGKKRPVEQSPQEPDSSSGIGTG	QQPAKRLNF-GQTG	DSESV 182
Db	123	RVLEPLGLVEE	GAKTAPGKKRPLIESPQQPDSSSGIGTG	KKQPAKRLVFEDE	TGADGP 182
QY	183	PDPQPLGEPP	ATPAVGPPTMASGGAPMADNNEGAD	GVGNASGNW	HCDSTWLGDRVITT 242
Db	183	PEGSTSG-----	AMSDSEMRRAAGAAVEGGGAGD	GVGNASGNW	HCDSTWSEGHVTTT 236
QY	243	STRTWALPT	YNNHLYKQISSASTGASNDNH	FGYSTPWGYF	EDENRFHCHESPRDQRLIN 302
Db	237	STRTWALPT	YNNHLYKRLGE---SLQSTNTYNG	FSTPWGYF	EDENRFHCHESPRDQRLIN 292



QY	303	NWNGFRPKRLNFKLFNIQVKEVTTNDGVTTLANNLSTVQVFSDESYQLPYVLGSAHQGC	3622
Db	293	NNWGNRPKAMRKVTFNIQVKEVTTNNGEITVANNLSTVQIFADSSYELPYMDAGQEGS	3522
QY	363	LPPEPADVEMIPQYGY--LTLNNGSQAAGRSSFYCLEYFPSSQMLRTGNNFTFSYTEEEV	419
Db	353	LPPEPNDVEMVPQYGYCGLVATGNTSQQQCTDBNAFYCLEYFPSSQMLRTGNNFEITYSFEKV	412
QY	420	PEHSSYAHQSQSLDRMLNPLIDQYLYLNRTON----QSGSAQNKDLLFSRGSBAGMSVQP	475
Db	413	PEHSMYAHQSQSLDRMLNPLIDQYLMGLQSTTGTTLNAGTATTN--FTKLRPTNFSNFK	469
QY	476	KNWLPGPCYRQQRVSKTKTDNNNSNFTWTGAS---KY---NLNGRESTINPGTAMASHK	528
Db	470	KNWLPGPSIKQQGFSKTA--NQNKYIPATGSDSLIKYETHSTLDGRWSALTPGPMMATAG	527
QY	529	DDEDKFFPMGCVMIFGKESAGASNTALDNVMITDEEIEIKATNPVATERPGTYAVNFOSSS	588
Db	528	PADSK-FSNSQLIFAGPKQNGNTATVPGTILFTSEELATNATDITDMGNLPGGDQSNS	586
QY	589	TDPATGDVHAMGALPGMWQDRDYYLQGPWAKIPHTDGHFHPSPPLMGFGGLKNPPQIL	648
Db	587	NLPYVDRLLTALGAVPGMWQNRDIYYQGPWAKIPHTDGHFHPSPPLIGFGGLKHPPOIF	646
QY	649	IKNTPVPANPPAEFSATKFAFITQYSTQGVSEIEMELQKENSKRWNPEVOYTSNYAKS	708
Db	647	IKNTPVPANPATTFSSSTPVNSFITQYSTQGVSVQIDWEIQERSKRWNPEVOFTSNYGQO	706
QY	709	ANVDFTVDNNGLYTEPRPIGTRYLTRPL	736
Db	707	NSLIWAPDAAGKYTEPRAIGTRYLTRPL	734

```

RESULT 2
US-09-532-594B-16
; Sequence 16, Application US/09532594B
; Patent No. 6468524
; GENERAL INFORMATION:
; APPLICANT: Chorini, John A.
; APPLICANT: Kofin, Robert M.
; APPLICANT: Safer, Brian
; APPLICANT: Davidson, Beverly
; TITLE OF INVENTION: AAV4 VECTOR AND USES THEREOF
; FILE REFERENCE: 14014.025202
; CURRENT APPLICATION NUMBER: US/09/532,594B
; CURRENT FILING DATE: 2000-03-22
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 598
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence; No. 6468524e =
; NAME/KEY: misc feature
; OTHER INFORMATION: AAV4 capsid protein VP2
US-09-532-594B-16

```

	Query Match	45.9%;	Score 1830.5;	DB 4;	Length 598;
	Best Local Similarity	57.8%;	Pred. No. 6,3e-152;		
	Matches 355,	Conservative 74;	Mismatches 154;	Indels 31;	Gaps 10;
QY	138 TAPGKKRPVEOSPOEPPDSSSGTGTGOQPAAKRLNF-GOTGSESVPDPQLGEPPATPA	:	:   :	:	
Db	1 TAPGKKRPLIESPQQPDSTGTGGKKGQPAKKLVFEDETGA GDGPPEGSTG-----A	:	:   :	:	
QY	197 AVGPPTMASGGAGAPMADNNEGADVGNASGNWCHDSTWLGDRIYTTSTRTWALPTYNNHL	:	:	:	:
Db	55 MSDDSEMRAAAGAAVEGGQGADGVGNASGDWHCDSTWSEGHVTTTSTRTWLPTYNNHL	:	:	:	:
QY	257 YKQISSASTGASNDNHFYGYSTPWGYFDENRFHCHFSPROWQRULINNNGFRPKRLNFKL	:::::	:	:	:

Db	115	YKRLGE----	SLQSNTYNGFSTPMWGYFDENRFHCHFSRDMQRLINNNGMRPKAMRVKI	170
Qy	317	FNIOYKEVTTNDGVTTIANNTLSTFVOVFS	DSEYQLPYVLGSAHQCLPEPADVEMIPOY	376
Db	171	FNIOYKEVTTNSGETTVANNLTSTVQIFADSS	YELPYMDAGQEGSLPPFENDVEMVPOY	230
Qy	377	GY----LTLNNGSQAVGRSSFYCLEYEP	SQMLRTGNNFTFSYTFEEVPEFHSYAHSQSLDR	433
Db	231	GYCGLVGTNTSQOQOTDRNAFYCLEYEP	SQMLRTGNNFEITYTSFEKVPFHSMYAHSQSLDR	290
Qy	434	LMNPLIDQYLYYLNRTON---	QSGSAONKDLLFSRGSAPGMSVOPKXWLPBCYRQORV	489
Db	291	LMNPLIDQYLMGLQSTTTGTLNAGTATTN--	FTKLRPTNFSNFKKXWLPGPSIKQOGF	347
Qy	490	SKTKTDNNNSNFTWTGAS--KY-----	NNGRESIINPGTAMASHKDEDEKFFPMSGVMI	542
Db	348	SKTY--NQNYKIPATGSDSLIKYETHSTL	DGRWSALTPGPPMATAGPADSK-FSNSQLIF	404
Qy	543	FKGESAGASNTALDNVMTIDDEEIKATNP	VATERFGTVAVNFQSSSTPDATGDVHAMGAL	602
Db	405	AGPKONGNTATVPGLIFTSEELATNATD	TDMWGNLPGGDQSNNSLPTVDRLLTALGAV	464
Qy	603	PGMWQDRDVYLQGPWAKIPHDTGHHPS	PLMGFGGLKNPPOILLKNPVPANPBAEF	662
Db	465	PGMWQNRDIYYQGPWAKIPHDTGHHPS	PLIGFGGLKHPPOIFIKNTPVPANPATTF	524
Qy	663	SATKFAFITQYSTQGVSEIEIEMELOK	ENSKRWNPVQYTSNYAKSANVDFTVDNNGLYT	722
Db	525	SSTPEVNSFITQYSTQGVSVQIDWEIO	KERSKRWNPVQFTSNYGGQNSLMLMAPDAAGKYT	584
Qy	723	EPRPIGTRYLTRPL	736	
Db	585	EPRPIGTRYLTRPL	598	

```

RESULT 3
US-09-532-594B-18
; Sequence 18, Application US/09532594B
; Patient No. 6468524
; GENERAL INFORMATION:
; APPLICANT: Chorini, John A.
; APPLICANT: Kotin, Robert M.
; APPLICANT: Safer, Brian
; APPLICANT: Davidson, Beverly
; TITLE OF INVENTION: AAV4 VECTOR AND USES THEREOF
; FILE REFERENCE: 14014.025202
; CURRENT APPLICATION NUMBER: US/09/532,594B
; CURRENT FILING DATE: 2000-03-22
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 544
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence; No. 6468524e =
; OTHER INFORMATION: synthetic construct
; NAME/KEY: misc feature
; OTHER INFORMATION: AAV4 capsid protein VP3
US-09-532-594B-18

```

[illegible]

QY 324 VTNDGVTTIANNLTSTVQVFSDESEYQLPYVLGSAHQGLPPFPADVFMIPQYGY---LT 380  
Db 124 VTTSGEETVYANNLTSTVQIFADSSYELPYWDAGQEGSLPPFPNDYFMVWPQYGYGGLVT 183  
QY 381 LNNGSQAVGRSSFYCLEYFSPSOMLRTGNFTFSYTFEEVPFHSYAHQSOLDRLMPLID 440  
Db 184 GNTSQOQFDRNNAFYCLEYFSPSOMLRTGNFEITYSFEKVPFHSYAHQSOLDRLMPLID 243  
QY 441 QYLYVLNRTQN----QSGSAQNKDLLFSRGSFAGMSVQPKNWLPGPCYRQQRVSKTKTDN 496  
Db 244 QYLMGLQSTTTGTTLNAGTATTN---FTYLRPTNFSNFKGNWLPGPSIKQOQFSKTA--N 298  
QY 497 NNSNFTWTGAS---KY----NLNGRESINPGTAMASHKDEDEKFFPMGVMIFGKESAG 549  
Db 299 QNKIPATGSDSLIKYETHSTLDGRWSALTPGPPMATAGPADSK-FSNSQLIPAGPKONG 357  
QY 550 ASNTALDNVMTDEEIKATNPVATERFGTVAVNFQSSSTDPAATGDVHAMGALPGMWOD 609  
Db 358 NTATVPGLTLFTSEELATNATDTDMWGNLPGDQSNLPTVDRLTALGAVPGMWQN 417  
QY 610 RDVYLOGPIWAKIPHTDGHFHPSPMLMGFGGLKNPPQILIKTTPVANPPEAFSATKFAS 669  
Db 418 RDIYYQGPFWAKIPHTDGHFHPSPMLGGGLKHPPOIFIKTTPVANPATFSTPVNS 477  
QY 670 FITQYSTQGVSEIEMELQKENSKRMPVQYTSNYAKSANVDFTVNNGLYTEPRPIGT 729  
Db 478 FITQYSTQGVSVQIDWEIQEKERSKRMPVQYTSNYGQNSLMAADAGKYTERPRAIGT 537  
QY 730 RYLTIRPL 736  
Db 538 RYLTIRHL 544

RESULT 4

US-08-856-841-22  
; Sequence 22, Application US/08856841  
; Patent No. 6274307  
; GENERAL INFORMATION:  
; APPLICANT: ERWIN SOUTSCHEK  
; APPLICANT: MANFRED MOTZ  
; TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES  
; TITLE OF INVENTION: OR POLYPEPTIDES FROM THE PARVOVIRUS B19  
; NUMBER OF SEQUENCES: 28  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROOKS HAIDT HAFNER & DELAHUNTY  
; STREET: 99 PARK AVENUE  
; CITY: NEW YORK  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10016  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" FLOPPY DISC  
; COMPUTER: AT&T - IBM COMPATIBLE  
; OPERATING SYSTEM: MS-DOS Version 6.2  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/856,841  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/214,658  
; FILING DATE: 16-MARCH-1994  
; APPLICATION NUMBER: US 07/917,096  
; FILING DATE: 4-AUGUST-1992  
; APPLICATION NUMBER: PCT/DE91/00106  
; FILING DATE: 8-FEBRUARY-1991  
; APPLICATION NUMBER: DE40038262  
; FILING DATE: 8-FEBRUARY-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: ROBINSON, WILLIAM R.  
; REGISTRATION NUMBER: 27,224  
; REFERENCE/DOCKET NUMBER: LKR-9222-A  
; TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 697-3355  
TELEFAX: (212) 557-5635  
TELEX: NONE  
INFORMATION FOR SEQ ID NO: 22:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 543  
TYPE: AMINO ACID  
TOPOLOGY: LINEAR  
MOLECULE TYPE:  
DESCRIPTION: PEPTIDE  
HYPOTHETICAL: N/A  
ANTI-SENSE: N/A  
FRAGMENT TYPE: INTERNAL  
ORIGINAL SOURCE: SERUM FROM PATIENT WITH ACUTE  
ORIGINAL SOURCE: INFECTION (ERYTHEMA INFECTIONUM)  
IMMEDIATE SOURCE: GENETICALLY ENGINEERED PEPTIDE  
POSITION IN GENOME: N/A  
FEATURE:  
NAME/KEY: N/A  
LOCATION: N/A  
IDENTIFICATION METHOD: amino acid analysis and  
IDENTIFICATION METHOD: mass spectrometry  
OTHER INFORMATION:  
PUBLICATION INFORMATION:  
AUTHORS: COSSART, Y.E.  
AUTHORS: FIELD, A.M.  
AUTHORS: CANT, B.  
AUTHORS: WIDDOWS, D.  
TITLE: PARVOVIRUS-LIKE PARTICLES IN HUMAN SERA  
JOURNAL: LANCET  
VOLUME: 1  
ISSUE:  
PAGES: 72 - 73  
DATE: 1975  
DOCUMENT NUMBER:  
FILING DATE:  
PUBLICATION DATE:  
RELEVANT RESIDUES IN SEQ ID NO: 22:  
US-08-856-841-22  
Query Match 12.0%; Score 479.5; DB 4; Length 543;  
Best Local Similarity 27.1%; Pred. No. 1.7e-33;  
Matches 141; Conservative 79; Mismatches 226; Indels 75; Gaps 16;  
QY 212 ADNNEGADGVGNAS--GNWHCDSTWLGDRVITSTRTWALPTYNNHLYKQISSASTGASN 269  
Db 16 AEAETGAGGGGNSVKSMSWSEGAFTSANSVTCTFSRQPLIPYDPEHHYKVFSPAASSCHN 75  
QY 270 D-----NHFGYSTPWGYFDENRFHCHFSPRDWRLINNNGFRPKRLNFKLFNI 319  
Db 76 ASGREAKVCTISPIMGYSTPWRYLDENALNLFSPLEFQHLIENYGSIAPDALTWTISEI 135  
QY 320 QVKEVT--TNDGVTTIANNLTSTVQVFSDESEYQLPYVLGSAHQGLPPFPADVFMIPQYGY 377  
Db 136 AVKQVTDKGTGGV-QVTDSTTGRLCMLVDHEYKYPYVLGGQDTLAPBLPIWVYFPQYA 194  
QY 378 YLTINN-GSQAVG-----RSSFYCLEYFSPSOMLRTGNFTFSYTFEEVPFHSYAH 427  
Db 195 YLTGVDVNTQGISGDSKGLASESAFYVLEHSSFQLLGTGTASMSYKFPVPPEENLEG 254  
QY 428 SQSLDRMLNPLIDQYLYVLNRTQNQSGSAQNKDLLFSRGSFAGMSVQPKNWLPGPCYRQ 487  
Db 255 SQHYEMYNPL--YGRRLGVDTLGGDPKFRSL-----THEDHAIQPNQFMGPPLVNSV 306  
QY 488 RVSKTKTDNNSNFTWTGASKYNLNGRESINPG-TAMASHKDEDEKFFPMGVMIFGKE 546  
Db 307 STKEGDSNTGAKALTLSTGTSQNTRISLRPGPVSQPYHHWDTDKYVTGINAISHGQT 366  
QY 547 SAGASNTALDNV-----MTDEEIKATNPVATERFGTVAVNFQSSSTDPAATGDVHAM 599  
Db 367 TYG---NAEDKEYQGVGRFNEKEQLKQLQGLNMHTY-----FPNKGTOQYTDQIE-R 416  
QY 600 GALPGWVQDRDVYLGPIWAKIPHTDGHFHPSP-PLMGFGGLKNPPQILIKTTPVANP 658

Db 417 PLMVGSVMNRALHESQLSKIPNLDDSFKTQFALGWLHQPPOI----- 465  
QY 659 PAEFSATKFASTOYSTGOVSVEIEMEL-QKENSKEWNP 698  
Db 466 -----FLKOYAVGIMVTMTFKLGPRKATGRWNPQ 495

RESULT 5

US-08-856-841-16  
; Sequence 16, Application US/08856841  
; Patent No. 6274307  
; GENERAL INFORMATION:  
; APPLICANT: ERWIN SOUTSCHEK  
; APPLICANT: MANFRED MOTZ  
; TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES  
; TITLE OF INVENTION: OR POLYPEPTIDES FROM THE PARVOVIRUS B19  
; NUMBER OF SEQUENCES: 28  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROOKS HAIDT HAFNER & DELAHUNTY  
; STREET: 99 PARK AVENUE  
; CITY: NEW YORK  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10016  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" FLOPPY DISC  
; COMPUTER: AT&T - IBM COMPATIBLE  
; OPERATING SYSTEM: MS-DOS Version 6.2  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/856,841  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/214,658  
; FILING DATE: 16-MARCH-1994  
; APPLICATION NUMBER: US 07/917,096  
; FILING DATE: 4-AUGUST-1992  
; APPLICATION NUMBER: PCT/DE91/00106  
; FILING DATE: 8-FEBRUARY-1991  
; APPLICATION NUMBER: DE40038262  
; FILING DATE: 8-FEBRUARY-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: ROBINSON, WILLIAM R.  
; REGISTRATION NUMBER: 27,224  
; REFERENCE/DOCKET NUMBER: LKR-9222-A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 697-3355  
; TELEFAX: (212) 557-5635  
; TELEX: NONE  
; INFORMATION FOR SEQ ID NO: 16:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 500  
; TYPE: AMINO ACID  
; TOPOLOGY: LINEAR  
; MOLECULE TYPE:  
; DESCRIPTION: PEPTIDE  
; HYPOTHETICAL: N/A  
; ANTI-SENSE: N/A  
; FRAGMENT TYPE: INTERNAL  
; ORIGINAL SOURCE: SERUM FROM PATIENT WITH ACUTE  
; ORIGINAL SOURCE: INFECTION (ERYTHEMA INFECTIONOSUM)  
; IMMEDIATE SOURCE: GENETICALLY ENGINEERED PEPTIDE  
; POSITION IN GENOME: N/A  
; FEATURE:  
; NAME/KEY: N/A  
; LOCATION: N/A  
; IDENTIFICATION METHOD: amino acid analysis and  
; IDENTIFICATION METHOD: mass spectrometry  
; OTHER INFORMATION:  
; PUBLICATION INFORMATION:  
; AUTHORS: COSSART, Y.E.

; AUTHORS: FIELD, A.M.  
; AUTHORS: CANT, B.  
; AUTHORS: WIDDOWS, D.  
; TITLE: PARVOVIRUS-LIKE PARTICLES IN HUMAN SERA  
; JOURNAL: LANCET  
; VOLUME: 1  
; ISSUE:  
; PAGES: 72 - 73  
; DATE: 1975  
; DOCUMENT NUMBER:  
; FILING DATE:  
; PUBLICATION DATE:  
; RELEVANT RESIDUES IN SEQ ID NO: 16:  
; PUBLICATION INFORMATION:  
; AUTHORS: MANIATIS, T.  
; AUTHORS: FRITSCH, E.F.  
; AUTHORS: SAMBROOK, J.  
; TITLE: MOLECULAR CLONING  
; JOURNAL: COLD SPRING HARBOR, NY  
; VOLUME:  
; ISSUE:  
; PAGES: 1982  
; DATE: 1982  
; DOCUMENT NUMBER:  
; FILING DATE:  
; PUBLICATION DATE:  
; RELEVANT RESIDUES IN SEQ ID NO: 16:  
; PUBLICATION INFORMATION:  
; AUTHORS: SMITH, D.B.  
; AUTHORS: JOHNSON, K.S.  
; TITLE: SINGLE STEP PURIFICATION OF POLYPEPTIDES  
; TITLE: EXPRESSED IN ESCHERICHIA COLI AS FUSIONS WITH  
; TITLE: GLUTATHIONE S. TRANSFERASE  
; JOURNAL: GENE  
; VOLUME:  
; ISSUE: 67  
; PAGES: 31 - 40  
; DATE: 1988  
; DOCUMENT NUMBER:  
; FILING DATE:  
; PUBLICATION DATE:  
; RELEVANT RESIDUES IN SEQ ID NO: 16:  
; US-08-856-841-16  
Query Match 10.8%; Score 430.5; DB 4; Length 500;  
Best Local Similarity 29.5%; Pred. No. 2.9e-29;  
Matches 122; Conservative 45; Mismatches 170; Indels 77; Gaps 10;  
QY 47 LPGYKYLGPFGNDKGEPPVNAADAALEHDKAYDQQLKAGDNFYLRYNHADAEPQERLQ 106  
Db 133 LPGTNYVGFENLQAGPPQSAVDAAIRIHDFRYSQALKGINPYTHWTVADELLKNIK 192  
QY 107 DTSFGNLGRAVFOAKKRVLEPLGLVEEAKTAPGKKRVEQSPQEPDSSSGIGTKGQP 166  
Db 193 ETGFOQVYDYF-----TLKGAAPVAHFQ-----GSLPEVP 225  
QY 167 AKKRLNFGQTGDSSEVPDPQPIGEPPTPAVAGPTTMASGGGAPMADNNEGADGVNASG 226  
Db 226 AYNA-----SEKYPSTMTSVNSAEASTGA-----GGG-----GNSVKS 258  
QY 227 NWHCDSTWLGDRVITTTSTRTWALPTVNNHLYKQISSASTGASND-----NHVFGY 276  
Db 259 NMSEGATFSANSVTCTFSRQFLPYDPEHHYKVFSPASSCHNASGKEAKVCTISPIMGY 318  
QY 277 STPMGYEDFNRFHCHFSRDMQRLINNMGFRPKRLNFKLNIQVKEVT--TNDGVTTIA 334  
Db 319 STPMRYLDENALNLFSPLEFQHLIENYGSIAIPDALVTIIEIAVKDVTDKTGGV-QVT 377  
QY 335 NNLTSTVOVFSDESYQLPYVUSAHQGLPPFPADVFMTPQYGYLTINN-GSQAVG----- 389  
Db 378 DSTGRLCMLVDHEYKYPYVIGOGQDTLAPLPIWVYFPPOYAYLTVGDVNTQGISGSK 437  
QY 390 -----RSSFYCLEYFPSSQMLRTGNNFTFSYTFEEVPPHSSYAHQSQSLRLMNP 438























GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 2003, 18:34:40 ; Search time 35.4414 Seconds  
(without alignments)  
2466.245 Million cell updates/sec

Title: US-09-807-802A-13  
Perfect score: 3989  
Sequence: 1 MAADGYLPDWLEDNLSEGIR.....NNGLYTEPRPIGTRYLTRPL 736

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:\*

1: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep1:\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep2:\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep3:\*  
13: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3402.5	85.3	735	15	US-10-038-972A-13 Sequence 13, Appl
2	3402.5	85.3	735	15	US-10-293-478-1 Sequence 1, Appl
3	2759.5	69.2	598	15	US-10-038-972A-14 Sequence 14, Appl
4	2481.5	62.2	533	15	US-10-038-972A-15 Sequence 15, Appl
5	1410	35.3	756	15	US-10-205-942-4 Sequence 4, Appl
6	119.5	3.0	1394	12	US-09-839-996-2 Sequence 2, Appl
7	119.5	3.0	1394	15	US-10-080-505-2 Sequence 2, Appl
8	118	3.0	2344	10	US-09-815-242-12713 Sequence 12713, A
9	117.5	2.9	1395	15	US-10-080-505-7 Sequence 7, Appl
10	117	2.9	1016	10	US-09-815-242-5845 Sequence 5845, Ap
11	116.5	2.9	5795	10	US-09-815-242-12610 Sequence 12610, A
12	115.5	2.9	410	15	US-10-022-809-2 Sequence 2, Appl
13	112	2.8	1441	15	US-10-170-682-3 Sequence 3, Appl
14	111.5	2.8	1265	11	US-09-862-027-80 Sequence 80, Appl
15	111	2.8	433	11	US-09-859-888-4 Sequence 4, Appl
16	111	2.8	2364	15	US-10-156-761-7834 Sequence 7834, Ap

17	110	2.8	938	15	US-10-029-217A-28	Sequence 28, Appl
18	110	2.8	938	15	US-10-029-217A-29	Sequence 29, Appl
19	109.5	2.7	1848	12	US-09-839-996-6	Sequence 6, Appl
20	109.5	2.7	1848	15	US-10-080-505-6	Sequence 6, Appl
21	109	2.7	1545	12	US-09-839-996-4	Sequence 4, Appl
22	109	2.7	1545	15	US-10-080-505-4	Sequence 4, Appl
23	108	2.7	1463	11	US-09-971-536-69	Sequence 69, Appl
24	106.5	2.7	331	11	US-09-996-561-1	Sequence 1, Appl
25	106.5	2.7	331	11	US-09-884-948-1	Sequence 5, Appl
26	106.5	2.7	331	15	US-10-112-488-5	Sequence 2, Appl
27	106.5	2.7	332	10	US-09-892-864A-2	Sequence 4, Appl
28	106.5	2.7	407	15	US-10-112-488-13	Sequence 13, Appl
29	106.5	2.7	407	15	US-10-112-488-13	Sequence 200, App
30	106.5	2.7	1023	14	US-10-137-866-200	Sequence 200, App
31	106.5	2.7	1023	14	US-10-146-726-200	Sequence 200, App
32	106.5	2.7	1023	14	US-10-146-727-200	Sequence 200, App
33	106.5	2.7	1023	14	US-10-146-788-200	Sequence 200, App
34	106.5	2.7	1023	14	US-10-152-380-200	Sequence 200, App
35	106.5	2.7	1023	14	US-10-153-934-200	Sequence 200, App
36	106.5	2.7	1023	15	US-10-028-072-200	Sequence 200, App
37	106.5	2.7	1023	15	US-10-121-049-200	Sequence 200, App
38	106.5	2.7	1023	15	US-10-123-904-200	Sequence 200, App
39	106.5	2.7	1023	15	US-10-140-470-200	Sequence 200, App
40	106.5	2.7	1023	15	US-10-175-746-200	Sequence 200, App
41	106.5	2.7	1023	15	US-10-176-918-200	Sequence 200, App
42	106.5	2.7	1023	15	US-10-176-921-200	Sequence 200, App
43	106.5	2.7	1023	15	US-10-137-865-200	Sequence 200, App
44	106.5	2.7	1023	15	US-10-140-474-200	Sequence 200, App
45	106.5	2.7	1023	15	US-10-142-431-200	Sequence 200, App

ALIGNMENTS

RESULT 1  
US-10-038-972A-13  
; Sequence 13, Application US/10038972A  
; Publication No. US20020192823A1  
; GENERAL INFORMATION:  
; APPLICANT: J. Bartlett  
; TITLE OF INVENTION: AAV VECTORS AND METHODS  
; FILE REFERENCE: 28335/36996US  
; CURRENT APPLICATION NUMBER: US/10/038, 972A  
; CURRENT FILING DATE: 2002-01-04  
; PRIOR APPLICATION NUMBER: US 60/260, 124  
; PRIOR FILING DATE: 2001-01-05  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 13  
; LENGTH: 735  
; TYPE: PRT  
; ORGANISM: adeno-associated virus 2 VP1 capsid protien  
; US-10-038-972A-13

Query Match	Best Local Similarity	Score	DB ID	Length	Matches	Conservative	Pred.	Mismatches	Indels	Gaps
QY	1	MAADGYLPDWLEDNLSEGI	60	60	60	60	60	60	60	60
DB	1	MAADGYLPDWLEDTLSEGI	60	60	60	60	60	60	60	60
QY	61	KGEVNAADAALHDKAYDQ	120	120	120	120	120	120	120	120
DB	61	KGEVNAADAALHDKAYDQ	120	120	120	120	120	120	120	120
QY	121	AKKRVLEPLGLVEEGAKT	180	180	180	180	180	180	180	180
DB	121	AKKRVLEPLGLVEEPVKTA	180	180	180	180	180	180	180	180
QY	181	SVDPDQPLGEPPATPAAV	240	240	240	240	240	240	240	240
DB	181	SVDPDQPLGQPPAAPSGL	240	240	240	240	240	240	240	240



```
Qy 241 TTSTRWALPTYNNHLYKQISSASTGASNDNHFGYSTPMGYFDENRFHCHFSPRDWORL 300
    |||||
Db 241 TTSTRWALPTYNNHLYKQISSOS -GASNDNHFGYSTPMGYFDENRFHCHFSPRDWORL 299
Qy 301 INNNWGFRPKRLNFKLFNIQVKEVTNDGVTTIANNLTSTVOVFSDEYQLPYVLGSAHQ 360
    |||||
Db 300 INNNWGFRPKRLNFKLFNIQVKEVTQNDGTTIANNLSTVQVFTDSEYQLPYVLGSAHQ 359
Qy 361 GCLPPFPADVFMIPQYGYLTLNNGSOAVGRSSFYCLEYFP SQMLRTGNFTFSYTFEEVP 420
    |||||
Db 360 GCLPPFPADVFMVPOYGYLTLNNGSOAVGRSSFYCLEYFP SQMLRTGNFTFSYTFEDVP 419
Qy 421 FHSSYAHQSOLDRLMNP LIDQYLYLNRTQNGSGSAQNKDLLFSRGS PAGMSVQPKNMLP 480
    |||||
Db 420 FHSSYAHQSOLDRLMNP LIDQYLYLSRTNTPSGTTQSRLOFSQAGASDIRDQSRNMLP 479
Qy 481 GPCYRQORVSKTKTDNNNSNFTWTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMGV 540
    |||||
Db 480 GPCYRQORVSKTSADNNNSEYSWTGATKYHLNGRDSLVPNGPAMASHKDEDEKFFPOSGV 539
Qy 541 MIFGESAGASNTALDNVMITDEEIKATNPVATERFGTVAVNFQSSSTD PATGDVHAMG 600
    |||||
Db 540 LIFGKGSEKTNVDIEKVMITDEEIRITNPVATEQYGSVSTNLQRGNQAATADVNTQG 599
Qy 601 ALPGMWQDRDLYLQGPIMAKI PHTDGHFHPSPLMGFGGLKNPPQILIKNTPV PANPA 660
    |||||
Db 600 VLPGMWQDRDLYLQGPIMAKI PHTDGHFHPSPLMGFGGLKHP PQLIKNTPV PANPST 659
Qy 661 EFSATKFAFITQYSTGQVSVEIEMELQKENS KRMNPEVOYTSNYAKSANVDFTVDNNGL 720
    |||||
Db 660 TFSAAKFAFITQYSTGQVSVEIEMELQKENS KRMNPEIQYTSNYNKS VNVDFTVDTNGV 719
Qy 721 YTEPRPIGTRYLTRPL 736
    |||||
Db 720 YSEPRPIGTRYLTRNL 735
```

## RESULT 2

```
US-10-293-478-1
; Sequence 1, Application US/10293478
; Publication No. US20030078411A1
; GENERAL INFORMATION:
; APPLICANT: PATEL, SALIL D.
; APPLICANT: MCARTHUR, JAMES G.
; TITLE OF INVENTION: ADENO-ASSOCIATED VIRUS CAPSID IMMUNOLOGIC DETERMINANTS
; FILE REFERENCE: 39672
; CURRENT APPLICATION NUMBER: US/10/293,478
; PRIOR FILING DATE: 2002-11-14
; PRIOR APPLICATION NUMBER: US/09/321,589
; PRIOR FILING DATE: 1999-05-28
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 735
; TYPE: PRT
; ORGANISM: Adeno-associated virus
US-10-293-478-1
```

Query Match 85.3%; Score 3402.5; DB 15; Length 735;  
Best Local Similarity 83.3%; Pred. No. 1.3e-281;  
Matches 613; Conservative 51; Mismatches 71; Indels 1; Gaps 1;

```
Qy 1 MAADGYLPDWLENLSEGIREWMDLKGAPKPKRANQOKODDGRGLVPGYKYLGPFGNGLD 60
    |||||
Db 1 MAADGYLPDWLENLSEGIRQMWKLKGP PPKPAERHKDSSRGLVLPGYKYLGPFGNGLD 60
Qy 61 KGEPVNAADAALAEHDKAYDQQLKAGDNPYLRYNHADAFAEQERLQEDTSFGN LGRAVFQ 120
    |||||
Db 61 KGEPVNEADAALAEHDKAYDROLDSGDNPYLRKYNHADAFAEQERLKEDETSFGN LGRAVFQ 120
Qy 121 AKKRVLEPLGLVEGAKTAPGKKRPVEQSPQEPDSSSGIGKTGOQPAKKRLNFGQTGDSE 180
    |||||
```

```
Db 121 AKKRVLEPLGLVEEPVKTAPEKKRPVEHSFVEPDSSSGTGKAGQOPARKRLNFGQTGDAD 180
Qy 181 SVPDPOLGEP PATPAVGPPTMASGGGAPMADNNEGADGVNASGNWCHDSTWLGD RVI 240
    |||||
Db 181 SVPDPOLGQPPAPASGLGTNTMATGSGAPMADNNEGADGVN SGNWCHDSTW MGDRVI 240
Qy 241 TTSTRWALPTYNNHLYKQISSASTGASNDNHFGYSTPMGYFDENRFHCHFSPRDWORL 300
    |||||
Db 241 TTSTRWALPTYNNHLYKQISSOS -GASNDNHFGYSTPMGYFDENRFHCHFSPRDWORL 299
Qy 301 INNNWGFRPKRLNFKLFNIQVKEVTNDGVTTIANNLTSTVOVFSDEYQLPYVLGSAHQ 360
    |||||
Db 300 INNNWGFRPKRLNFKLFNIQVKEVTQNDGTTIANNLSTVQVFTDSEYQLPYVLGSAHQ 359
Qy 361 GCLPPFPADVFMIPQYGYLTLNNGSOAVGRSSFYCLEYFP SQMLRTGNFTFSYTFEEVP 420
    |||||
Db 360 GCLPPFPADVFMVPOYGYLTLNNGSOAVGRSSFYCLEYFP SQMLRTGNFTFSYTFEDVP 419
Qy 421 FHSSYAHQSOLDRLMNP LIDQYLYLNRTQNGSGSAQNKDLLFSRGS PAGMSVQPKNMLP 480
    |||||
Db 420 FHSSYAHQSOLDRLMNP LIDQYLYLSRTNTPSGTTQSRLOFSQAGASDIRDQSRNMLP 479
Qy 481 GPCYRQORVSKTKTDNNNSNFTWTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMGV 540
    |||||
Db 480 GPCYRQORVSKTSADNNNSEYSWTGATKYHLNGRDSLVPNGPAMASHKDEDEKFFPOSGV 539
Qy 541 MIFGESAGASNTALDNVMITDEEIKATNPVATERFGTVAVNFQSSSTD PATGDVHAMG 600
    |||||
Db 540 LIFGKGSEKTNVDIEKVMITDEEIRITNPVATEQYGSVSTNLQRGNQAATADVNTQG 599
Qy 601 ALPGMWQDRDLYLQGPIMAKI PHTDGHFHPSPLMGFGGLKNPPQILIKNTPV PANPA 660
    |||||
Db 600 VLPGMWQDRDLYLQGPIMAKI PHTDGHFHPSPLMGFGGLKHP PQLIKNTPV PANPST 659
Qy 661 EFSATKFAFITQYSTGQVSVEIEMELQKENS KRMNPEVOYTSNYAKSANVDFTVDNNGL 720
    |||||
Db 660 TFSAAKFAFITQYSTGQVSVEIEMELQKENS KRMNPEIQYTSNYNKS VNVDFTVDTNGV 719
Qy 721 YTEPRPIGTRYLTRPL 736
    |||||
Db 720 YSEPRPIGTRYLTRNL 735
```

## RESULT 3

```
US-10-038-972A-14
; Sequence 14, Application US/10038972A
; Publication No. US20020192823A1
; GENERAL INFORMATION:
; APPLICANT: J. Bartlett
; TITLE OF INVENTION: AAV VECTORS AND METHODS
; FILE REFERENCE: 28335/36996US
; CURRENT APPLICATION NUMBER: US/10/038,972A
; CURRENT FILING DATE: 2002-01-04
; PRIOR APPLICATION NUMBER: US 60/260,124
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
; LENGTH: 598
; TYPE: PRT
; ORGANISM: adeno-associated virus 2 VP2 capsid protien
US-10-038-972A-14
```

Query Match 69.2%; Score 2759.5; DB 15; Length 598;  
Best Local Similarity 82.8%; Pred. No. 7.5e-227;  
Matches 495; Conservative 43; Mismatches 59; Indels 1; Gaps 1;

```
Qy 139 APGKKRPVEQSPQEPDSSSGIGKTGOQPAKKRLNFGQTGDSVDPQPLGEP PATPAV 198
    |||||
Db 2 APGKKRPVEHSFVEPDSSSGTGKAGQOPARKRLNFGQTGDADVDPQPLGQPPAAPSGL 61
Qy 199 GPTMASGCGAPMADNNEGADGVNASGNWCHDSTWLGD RVI TTSTRWALPTYNNHLYK 258
    |||||
```

```
Db      62 GTNTMATGSGAPMADNNEGADGVNSSGNWHCDSTWMGDRVITTTSTRTWALPTYNNHLTK 121
QY      259 QISSASTGASNDNHFGYSTPMGYFDNFRFHCHFSPRDWORLINNMGFRPKRLNFKLEN 318
Db      122 QISSQS-GASNDNHFGYSTPMGYFDNFRFHCHFSPRDWORLINNMGFRPKRLNFKLEN 180
QY      319 IOYKEVTNDGVTTIANNLTSTVQVFSDEYQLPYVLGSAHOGCLPPFPADVEMIPQYGY 378
Db      181 IOYKEVTQNDGTTTIANNLSTVQVFTDSEYQLPYVLGSAHOGCLPPFPADVEMVPQYGY 240
QY      379 LTIANGSQAVGRSSFYCLEYFPSSQMLRTGNNFTFSYTFEEVPHSSYAHQSOLDRLMNP 438
Db      241 LTIANGSQAVGRSSFYCLEYFPSSQMLRTGNNFTFSYTFEDVPHSSYAHQSOLDRLMNP 300
QY      439 IDOYLTYLNRTONQSGSAQNKLDFSRCSPAGMSVQPKNWLPGPCYRQQRVSKTKTDNN 498
Db      301 IDOYLTYLNRTPSGTTTQSRLOFSQAGASDIRDQSRNWLPGPCYRQQRVSKTSADNN 360
QY      499 SNFTWTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMGCVMI FGKESAGASNTALDNV 558
Db      361 SEYSWTGATKYHLNGRDSLVPNGPRAMASHKDEBEKFFPQSGVLI FGKQSEKTNVDIEKV 420
QY      559 MITDEEIKATNPVATERFGTVAVNFQSSSTD PATGDVHAMGALPGMWQDRDVTYLOGP 618
Db      421 MITDEEIRITNPVATEQYGSVSTNLQRGNRQAATADVNTQVLPGMVWQDRDVTYLOGP 480
QY      619 WAKIPHTDGHFHPSPMLMGFGGLKNPPOILIKNTPVPANPPAEFSATKFAFITQYSTGQ 678
Db      481 WAKIPHTDGHFHPSPMLMGFGGLKHPPOILIKNTPVPANPSTTFSAAKFASFITQYSTGQ 540
QY      679 VSVETIEMELQENSKRWNPVQYTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 736
Db      541 VSVETIEMELQENSKRWNPETQYTSNYNKSXVNDFTVDINGVSEPRPIGTRYLTRNL 598
```

```
RESULT 4
US-10-038-972A-15
; Sequence 15, Application US/10038972A
; Publication No. US20020192823A1
; GENERAL INFORMATION:
; APPLICANT: J. Bartlett
; TITLE OF INVENTION: AAV VECTORS AND METHODS
; FILE REFERENCE: 28335/36996US
; CURRENT APPLICATION NUMBER: US/10/038, 972A
; CURRENT FILING DATE: 2002-01-04
; PRIOR APPLICATION NUMBER: US 60/260,124
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 15
; LENGTH: 533
; TYPE: PRT
; ORGANISM: adeno-associated virus 2 VP3 capsid protien
US-10-038-972A-15
```

```
Query Match      62.2%; Score 2481.5; DB 15; Length 533;
Best Local Similarity 83.3%; Pred. No. 3.4e-203;
Matches 445; Conservative 37; Mismatches 51; Indels 1; Gaps 1;

QY      203 MASGGAPMADNNEGADGVNASGNWHCDSTWLGDRVITTTSTRTWALPTYNNHLTKQISS 262
Db      1 MATGSGAPMADNNEGADGVNSSGNWHCDSTWMDRVITTTSTRTWALPTYNNHLTKQISS 60
QY      263 ASTGASNDNHFGYSTPMGYFDNFRFHCHFSPRDWORLINNMGFRPKRLNFKLENIOVK 322
Db      61 QS-GASNDNHFGYSTPMGYFDNFRFHCHFSPRDWORLINNMGFRPKRLNFKLENIOVK 119
QY      323 EYTTNDGVTTIANNLTSTVQVFSDEYQLPYVLGSAHOGCLPPFPADVEMIPQYGYLTN 382
Db      120 EYTONDGTITIANNLSTVQVFTDSEYQLPYVLGSAHOGCLPPFPADVEMVPQYGYLTN 179
QY      383 NGSQAVGRSSFYCLEYFPSSQMLRTGNNFTFSYTFEEVPHSSYAHQSOLDRLMNPIDQY 442
```

```
Db      180 NGSQAVGRSSFYCLEYFPSSQMLRTGNNFTFSYTFEDVPHSSYAHQSOLDRLMNPIDQY 239
QY      443 LYILNRTONQSGSAQNKLDFSRCSPAGMSVQPKNWLPGPCYRQQRVSKTKTDNNNSNFT 502
Db      240 LYILSRNTTPSGTTTQSRLOFSQAGASDIRDQSRNWLPGPCYRQQRVSKTSADNNNSEYS 299
QY      503 WTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMGCVMI FGKESAGASNTALDNVMTD 562
Db      300 WTGATKYHLNGRDSLVPNGPRAMASHKDEBEKFFPQSGVLI FGKQSEKTNVDIEKMTTD 359
QY      563 EEEIKATNPVATERFGTVAVNFQSSSTD PATGDVHAMGALPGMWQDRDVTYLOGPIMAKI 622
Db      360 EEEIRITNPVATEQYGSVSTNLQRGNRQAATADVNTQVLPGMVWQDRDVTYLOGPIMAKI 419
QY      623 PHTDGHFHPSPMLMGFGGLKNPPOILIKNTPVPANPPAEFSATKFAFITQYSTGQVSE 682
Db      420 PHTDGHFHPSPMLMGFGGLKHPPOILIKNTPVPANPSTTFSAAKFASFITQYSTGQVSE 479
QY      683 IEMELQENSKRWNPVQYTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 736
Db      480 IEMELQENSKRWNPETQYTSNYNKSXVNDFTVDINGVSEPRPIGTRYLTRNL 533
```

```
RESULT 5
US-10-205-942-4
; Sequence 4, Application US/10205942
; Publication No. US20030053990A1
; GENERAL INFORMATION:
; APPLICANT: University of No. US20030053990A1th Carolina-Chapel Hill
; APPLICANT: Rabinowitz, Joseph E.
; APPLICANT: Samulski, Richard J.
; APPLICANT: Xiao, Weidong
; TITLE OF INVENTION: VIRUS VECTORS AND METHOD OF MAKING AND ADMINISTERING THE SAME
; FILE REFERENCE: 5470-186
; CURRENT APPLICATION NUMBER: US/10/205, 942
; CURRENT FILING DATE: 2002-07-26
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 756
; TYPE: PRT
; ORGANISM: Adeno-associated virus
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(2271)
; OTHER INFORMATION: B19/AAV chimeric capsid coding sequence
US-10-205-942-4
```

```
Query Match      35.3%; Score 1410; DB 15; Length 756;
Best Local Similarity 42.7%; Pred. No. 1.8e-111;
Matches 312; Conservative 95; Mismatches 269; Indels 54; Gaps 16;

QY      1 MAADGYLPDWLEDNLTSEGIREWMDLKPAGAPKPKANQKODDGRGLVLPGYKYLGPENG 60
Db      1 MAADGYLPDWLEDTLSEGIROWWKLKPGPPPKPAERHKDSDRGLVLPGYKYLGPENG 60
QY      61 KGEPVNADAAALAHDKAYDQOLKAGDNPYLRYNHADAERLQEDTSGFGLGRAVFQ 120
Db      61 KGEPVNEADAAALAHDKAYDRQLDSGDNPLYLRYNHADAERLQEDTSGFGLGRAVFQ 120
QY      121 AKKRVLEPLGLVEEGAKTAPGKKRPVEQSPQEPDSSSGIGTKGOQPAKKRLNFGQTGDSE 180
Db      121 AKKRVLEPLGLVEEPVKTAPEGKRVEHSVPEPDSSSGTGKAGQOPARKRLNFGQTGDAD 180
QY      181 SVDPDQPLGEPPTPAVGPPTMASGGGAPMADNNEGADGVNASGNWHCDSTWLGDRVI 240
Db      181 SVDPDQPLGQPPAAPSGLGTINTWTSVNSAE-ASTGAGGGGSSNVKSMSEGATFSANSVT 239
QY      241 TTSTRTWALPTYNNHLTKQISSASTGASND-----NHFGYSTPMGYFDNFRFHCH 290
Db      240 CTFSRQFLIPYDPEHHYKVFSPAASSCHNASGKEAKVCTTISPIMGYSTPWRXYLDENALNL 299
QY      291 HFSPRDWORLINNMGFRPKRLNFKLENIOYKEVT--TNDGVTTIANNLTSTVQVFSDE 348
```

Db 300 FFSPLFEQHLIENYGSIAPDALTVTITSEIAVKDVTDKTGGV-QVTDSTGRCLMLVDHE 358  
Qy 349 YOLPYVLGSAHQGLPPFPADVFMIPQYGLTLNN-GSQAVG-----RSSFYCLEY 398  
Db 359 YKYPYVLGGQDGLAPLPIWVFPFQYAYLVGDVNTQGISGDSKLAESSEAFVLEH 418  
Qy 399 FPSQMLRTGNNTFSYTFEEVFPFHSSYAHOSLDRLMPLIDQYLYLNRTQNOSSAQN 458  
Db 419 SSFQLGTGTATMSYKFPVPPELNEGCSQHFYEMYNPL--YGSRLGVPTLGGDPKF 475  
Qy 459 KDLFSRGSFAGMSVQPKNWLPGPCYRQQRVSKTKTDNNNSNFTWTGASKYNLNGRESII 518  
Db 476 RSL-----THEDHAIQPNFMFGLVNSVSTKEGDSNTGAGKALTGSTGTSQNTRI SL 530  
Qy 519 NPG-TAMASHKDEDEKFFPMSGVMI FGKESAGASNTALDNV-----MITDEEIKATN 570  
Db 531 RGPVVSQPHHMDTKVVTGINAISHGQTYG--NADKEYQQGVGRFPNEKEQLKOLQ 587  
Qy 571 PVATERFGYAVNFQSSSTDPATGVDHAMGALLPGMWQDRDVLQGPWAKIPHTDGHF 630  
Db 588 GLNMHTY-----FPNKGTQOYTDQIE-RPLWGVSVNRRALHYESQLMSKIPNLDSFK 640  
Qy 631 PS-PLMGFGIKNPPQILIKNTVPANPPAEFSATKFPASFITQYSTGOVSVEIEMEL-Q 688  
Db 641 TQFAALGGWGHQPPQIFLK--ILPQSGPIGSIKSMGITTLVQYAVGIMVTMTFKLGP 698  
Qy 689 KENSKRWNP 698  
Db 699 RKATGRWNPQ 708

RESULT 6

US-09-839-996-2  
; Sequence 2, Application US/09839996  
; Publication No. US20030009010A1  
; GENERAL INFORMATION:  
; APPLICANT: St. Geme III, Joseph W.  
; Falkow, Stanley  
; TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentln Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/839, 996  
; FILING DATE: 20-Apr-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/296, 791  
; FILING DATE: 25-AUG-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Trecartin, Richard F.  
; REGISTRATION NUMBER: 31,801  
; REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249  
; TELEX: 910 277299  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1394 amino acids  
; TYPE: amino acid

TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-09-839-996-2

Query Match 3.0%; Score 119.5; DB 12; Length 1394;  
Best Local Similarity 17.4%; Pred. No. 0.69;  
Matches 152; Conservative 96; Mismatches 287; Indels 337; Gaps 38;

Qy 49 GYKYLQ-----PFNGL--DKGEPVNADAALAEHDKAYDQQLKAGDNPY---- 90  
Db 217 GYSYLGQDVRKAGEYGPLPIAGSKDGSMPFIYDA--EKQKWLINGILREGNPFEGKE 273  
Qy 91 -----LRYNHADAERLQEDTSFGNLRGAVF-----QAKRVLPLGLVEEGAKTA 139  
Db 274 NGFQVLRKSYFDEIFERDLH--TSLYTRAGNGVYTTISGNDGQGSITQKSGIPSEIKITL 331  
Qy 140 PGKKRPVEQSPOEPDSSSGIGTKQQQPAKRLNFGQT--GDSESVDPQPLGEPPATPA 196  
Db 332 ANMSLPL---KEKDKVHNPRYDGPNIYSPRLNGETLYFMDQKQ----- 372  
Qy 197 AVGPPTMASGGGAPMADNNEGADV--GNASGNMHCDSW--LGDRVITSTRTW---- 247  
Db 373 --GSLIFAS-----DINQAGGLYFEGNFTVSPNSQITWOGAGIHVSENSTVTKVNG 423  
Qy 248 -----ALPTYNNHLYKQI--SSASTG 266  
Db 424 VEHDRLSKIGKGLHVOAKENKGSISVGDKVILEQQAADQGNKQAFSEIGLVSGRGIV 483  
Qy 267 ASNDNHVFGYSTPWGYFDENRPFCHSPRDWQRLINNNGFRPKRLNFKLNIQVKEV-T 325  
Db 484 QLNDDK-----QFTDIDKEY-----FGFRGRLDLNGHSLTFKRIQN 519  
Qy 326 TNDGVTTIANLITSTVQVFSDEYQLPYLGSAGQCLPPPADVFMIPQYGLTLNNGS 385  
Db 520 TDEGAMLVNHTTQAANV-----TI-TGN 542  
Qy 386 QAVGRSSFYCLEYFPSQMLRTGNNTFSYTFEEVFPFHSSYA-----HSQSLDRMLNPLI 439  
Db 543 EST-----VLPNGNININKLDYRKEIAYNGWFGETDKNKHNGRLNIYKFTT 588  
Qy 440 DQYLYLNRTQNOGS-AQNKDLF-----SRGSPAGMSVQPKNWL 479  
Db 589 EDRTLISGNTLKGDIQTQKGLFSGRPTPHAYNHLNKRSEMEGIPQGEIIVMDHWI 648  
Qy 480 -----PGPCYRQQRVSKT--TDNNNSNFT-----WTGAS- 507  
Db 649 NRTFKAENFOIKGSAAVSRNVSSIEGNWTVSNANATFGVVPNQNTICTRSDWTGLTT 708  
Qy 508 -----KYNLNGRESIINPGTA-----MASHKDEDEKFFPMS 538  
Db 709 CQKVDLDTKVINSIPKTIQINGSINLTDNATANVKGLAKINGNVTLTNHSQ-----FTLS 763  
Qy 539 -----GWMIFGESAGASNTALD-NVMITDEEIKATNPVATERFGTVAVNFQSSSTD 590  
Db 764 NNATQIGNIRLSDNSTATVDNANLNGNVHLTDSAQFSLKN-----SHFSH 808  
Qy 591 PATGDVHAMGALLPGMWV-QDRDVLQGPWAKIPHTDGHFHPSPLMGGFGLKNPPQILI 649  
Db 809 QIQGDKGTTVLTENATWMTPSDTTLQNLTLNNSITTLNSAYSA-----SSNTPRRRSL 862  
Qy 650 KNPVNPANPPAEFSAT-----KFASFITQYSTGOVSV-----EIEWELQKENS 693  
Db 863 ETEETPTSAREHFNLTITVNGKLSGQGTFOFTSSLFGYKSDKLKLSNDAEGDYLISVRNTG 922  
Qy 694 RWNPE-----VOYTSNYAKSANVDFTVDN 718  
Db 923 K-EPETLEQLTLVESKDNQPLSDKIKFTLEND 953

RESULT 7

US-10-080-505-2  
; Sequence 2, Application US/10080505



```

; Publication No. US20030073166A1
; GENERAL INFORMATION:
; APPLICANT: St. Geme, Joseph W.
; TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS
; FILE REFERENCE: A-59941-1/RT/DCF/DHR
; CURRENT APPLICATION NUMBER: US/10/080, 505
; CURRENT FILING DATE: 2002-02-22
; PRIOR APPLICATION NUMBER: US 08/296, 791
; PRIOR FILING DATE: 1994-10-25
; PRIOR APPLICATION NUMBER: US 09/839, 996
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 1394
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
; US-10-080-505-2
```

```

Query Match          3.0%; Score 119.5; DB 15; Length 1394;
Best Local Similarity 17.4%; Pred. No. 0.69;
Matches 152; Conservative 96; Mismatches 287; Indels 337; Gaps 38;
```

```

QY 49 GYKYL-----PFGNL-DKEPVNADAALAHDAVDQQLKAGDNPY---- 90
Db 217 GYSYLGDDVRKAGEYGPLPIAGSKDGSMPFIYDA--EKQWMLINGILREGNPFEGKE 273
QY 91 -----LRYNHADAERFOERLOEDTSFGNIGRAVF-----QAKKRVLEPLGLVEEGAKTA 139
Db 274 NGFQVRSYFDEIFERDLH--TSLYTRAGNGVYTTISGNDGQSITQKSGIPSEIKITL 331
QY 140 PGKRPVEQSPOEPDSSSGIGTKQOPAKKRLNFGQT--GDSSEVPDPQPLGEPATPA 196
Db 332 ANMSLPL-----KEKDKVHNPRYDPNITYSPRLNGETLYFMDQKQ----- 372
QY 197 AVGPTMASGGGAPMADNNEGADV---GNASGNMHCSTW--LGDRVITTTSTRTW---- 247
Db 373 --GSLIFAS-----DINGAGGLYFEGNFTVSPNSQOTWQAGIHVSENSTVTWKVNG 423
QY 248 -----ALPTYNNHLYKQI--SSASTG 266
Db 424 VEHDRLSKIGKGLTHVQAKGENKGSISVGDGKVLLEQADQGNKQAFSEIGLVSGRGTV 483
QY 267 ASNDNHVEGYSTPWGYFDENRFHCHSPRDWQRLINNNGFPRKRLNFKLFNIQVKEV-T 325
Db 484 QLNDDK-----QFDTDKFY-----FGFRGRLDLNGHSLTFKRIQN 519
QY 326 TNDGVTTIANNLSTVQVFSDEYQLPYVLGSAHQGLPPPADVFEMIPOXYGLTLNNGS 385
Db 520 TDEGAMLVNHNTTQANV-----TI-TGN 542
QY 386 QAVGRSSFCLEYFSPQMLRTGNNFTFSYTFEEVPHSSYA-----HSQSLDRLMNPLI 439
Db 543 ESI-----VLPNGNNINIKLDYRKEIAYNGWFGETDKNKHNGRLNLIYKPTT 588
QY 440 DQYLYTLNQTONGSGS-AQNKDLF-----SRGSPAGMSVQPKMWL 479
Db 589 EDRITLLSGGTNLKGDITQTKGKLFPSGRPTPHAYNHLNKRWSEMEGIPQGEIWMHDWI 648
QY 480 -----PGPCYRQQRVSKTK---TDNNSNSFT-----WTGAS- 507
Db 649 NRTFKAENFOIKGSAVSRVSSIEGNWTVSNANATFGVVPNQONTICTRSDWTGLTT 708
QY 508 -----KYNLNGRESIINGTA-----MASHKDEDEKFFPMS 538
Db 709 CQKVDLPTKVINSLPKTQINGSINLTDNATANVKGLAKLNGNVTLTNHSQ-----FTLS 763
QY 539 -----GVNIFGKESAGASNTALD-NVMITDEEIKATNPVATERFGTVAVNFQSSSTD 590
Db 764 NNATQIGNIRLSDNSTATVDNANLNGNVLJDSAQFSLKN-----SHFSH 808
QY 591 PATGDVHAMGALPGMWV-QDRDVYLGPIWAKIPIHTDGHFHSPLMGFGGLKNPPQILI 649
```

```

Db 809 QIQGDKGTTVTLEENATWMPSDTTLQNLTINNSTITLNSAYSAA-----SSNTPRRRSL 862
QY 650 KNPVPANPPAEFSAT-----KFASFTQYSTQVSV----EIEWELOKENS 693
Db 863 EFTETPTSAEHRFNTLTVNGKLSGQGTFOFTSSLFGYKSDKLLSNDAGDYILSVNTG 922
QY 694 RMNPE-----VQYTSNVAKSANDFTVDNN 718
Db 923 K-EPETLEQLTLVESKDNQPLSDKLFTELEND 953
```

```

RESULT 8
US-09-815-242-12713
; Sequence 12713, Application US/09815242
; Patent No. US20020061569A1
```

```

; GENERAL INFORMATION:
```

```

; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
```

```

; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
```

```

; CURRENT APPLICATION NUMBER: US/09/815,242
```

```

; PRIOR FILING DATE: 2001-03-21
```

```

; PRIOR APPLICATION NUMBER: 60/191,078
```

```

; PRIOR FILING DATE: 2000-03-21
```

```

; PRIOR APPLICATION NUMBER: 60/206,848
```

```

; PRIOR FILING DATE: 2000-05-23
```

```

; PRIOR APPLICATION NUMBER: 60/207,727
```

```

; PRIOR FILING DATE: 2000-05-26
```

```

; PRIOR APPLICATION NUMBER: 60/242,578
```

```

; PRIOR FILING DATE: 2000-10-23
```

```

; PRIOR APPLICATION NUMBER: 60/253,625
```

```

; PRIOR FILING DATE: 2000-11-27
```

```

; PRIOR APPLICATION NUMBER: 60/257,931
```

```

; PRIOR FILING DATE: 2000-12-22
```

```

; PRIOR APPLICATION NUMBER: 60/269,308
```

```

; PRIOR FILING DATE: 2001-02-16
```

```

; NUMBER OF SEQ ID NOS: 14110
```

```

; SOFTWARE: FastSeq for Windows Version 4.0
```

```

; SEQ ID NO 12713
```

```

; LENGTH: 2344
```

```

; TYPE: PRT
```

```

; ORGANISM: Staphylococcus aureus
```

```

; US-09-815-242-12713
```

```

Query Match          3.0%; Score 118; DB 10; Length 2344;
Best Local Similarity 20.4%; Pred. No. 2;
Matches 87; Conservative 57; Mismatches 182; Indels 100; Gaps 18;
```

```

QY 174 GQTDSESV-PDQPLGEPATPAVG-----PTMASGGGAPMA 212
Db 326 GNGDGI GFAPSPGLGETLNGAAGVIGGLSNAPGFKLDTYHNTSTPNSSAKAKADPSN 385
QY 213 DNEGADGVGNASGNMHCSTWLGDRVITTTSTRTWALPTYNNHLYKQISSASTGASNDH 272
Db 386 VAGGAFAGFAVTTDSYGVASTYSSSTADNAAKLVQPTNN-----TFQDFDIN 434
QY 273 YFGYSTPWGYFDENRFHCHSPRDWQRLINNNGFPRKRLNFKLFNIQVKEVTTNDGVT 332
Db 435 YNG-----DTKVMTVKXAGQWTTRNI-SDWIAKSGTTNPSL-----SMTASTGATN 480
QY 333 IANNLSTVQVFSDEYQLPYVLGSAHQGLPP--FPADVEMIPOXYGLTLNNGSQAVGR 390
Db 481 LQOVQFTEFEYTESAVTVQVRYVDVTTGKDIIIPKTVSGVNDQV-----VTIDNQSALTA 535
QY 391 SSF-----YCLEYFSP-----QMLRTGNNFTFSYTFEEVPHSSYAHSQSLDRLMNPL 438
```

Db 536 KGNYTSDVSASTYNDTNKTKMTNAGSVTYFTDVKAPTVVNGTIEVGKTMNPI 595  
Qy 439 IDQYLYLNRTQNSGSAQNKDLFSGRSPAGMSV-QPKNWLPGPCYR--QQRVSKTKTD 495  
Db 596 V-----LTTDNGTGTNT---VTGLPSGLSYDSATNSIIGTPTKIGOSTVTIVSTD 645  
Qy 496 --NNSNFTWTGASKYNLNGRESIIN-----PGTAMASHKDEDEKFFPMGVMIFGES 547  
Db 646 QANNKSTTFT-----INVVDTPATVTPIGDKSE-VFSPISPINIATQDN 691  
Qy 548 AGASNT 553  
Db 692 SGNAV 697

RESULT 9  
US-10-080-505-7  
; Sequence 7, Application US/10080505  
; Publication No. US20030073166A1  
; GENERAL INFORMATION:  
; APPLICANT: St. Geme, Joseph W.  
; TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS  
; FILE REFERENCE: A-59941-1/RET/DCF/DHR  
; CURRENT APPLICATION NUMBER: US/10/080,505  
; PRIOR FILING DATE: 2002-02-22  
; PRIOR APPLICATION NUMBER: US 08/296,791  
; PRIOR FILING DATE: 1994-10-25  
; PRIOR APPLICATION NUMBER: US 09/839,996  
; PRIOR FILING DATE: 2001-04-20  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 7  
; LENGTH: 1395  
; TYPE: PRT  
; ORGANISM: Haemophilus influenzae  
US-10-080-505-7

Query Match 2.9%; Score 117.5; DB 15; Length 1395;  
Best Local Similarity 17.9%; Pred. No. 1;  
Matches 125; Conservative 73; Mismatches 211; Indels 291; Gaps 31;  
Qy 49 GYKYL-----PFGNL-DKGPVNAADAALEHDKAYDQOLKAGDNPY---- 90  
Db 217 GSYLGGDVRAKAGYGPLPIAGSKDGSFMTYDA--EKQKWLINGILREGNPFEGKE 273  
Qy 91 -----LRYNHADAEFOERLQEDTSFGGNLGRAV-----QAKKRVLEPIGLVEEGAKTA 139  
Db 274 NGFQVLRKSYFDEIFERDLH--TSLYTRAGNGVYITISGNDNGQSGITQSGIPSEIKITL 331  
Qy 140 PGKRPVEQSPQEPDSSSGIGKTGOQPAKKRLNFGQT--GDSSEVPDPQPLGEPATPA 196  
Db 332 ANMSLPL---KEKQVHNPRYDGNITYSPLNNGETLYEMDQK----- 372  
Qy 197 AVGPTTMASSGGA PMADNNEGADV--GNASGNWHCDSTW--LGDRVITTTSTRTW---- 247  
Db 373 --GSLIFAS-----DINQAGAGLYFEGNFTVSPNSNQTWQAGAGIHVSENSTVTWKVNG 423  
Qy 248 -----ALPYNNHLYKQI---SSASTG 266  
Db 424 VEHRLSKIGKTIHVQAKGENKGSISVGDGKYLLEQADDOGNKQAPSEIGLVSGRGTV 483  
Qy 267 ASNDNHFGYSTPWGYFDENRHFCHFSPRDWORLNNNWGFPRPKRLNKLFINIQVEV-T 325  
Db 484 QLNDDK-----QFDTDKFY-----FGFRGRLDLNGHSLTFKRIQN 519  
Qy 326 TNDGVTIANNLSTVQVFSDEYQLPYVLGSAHQGCLPPFPADVEMIPQYGYLTLLNGS 385  
Db 520 TDEGAMIVNHTTQAANY-----TI-TGN 542  
Qy 386 QAVGRSSFYCLEFSPQMLRTGNNTFTSYTFEEVPHSSYA-----HSQSLDRIMNPLI 439  
Db 543 ESI-----VLPNGNNINKLDYRKEIAVNGWFGJTDKNKNKHGRLNLIIYKPTT 588

Qy 440 DQYLYLNRTQNSGS-AQNKDLF-----SRGSPAGMSVQPKNWL 479  
Db 589 EDRTLILSGTNLKGDIQTQTKGLFFSGRPTPHAYNHLNKRWSEMEGIPQGEIYWDHWT 648  
Qy 480 -----PGPCYRQQRVSKTK---TDNNSNFT-----WTGAS- 507  
Db 649 NRTFKAENFQIKGSAAVSRVNSIIEGNMTVSNANATFGVYPNQNTICTRSDWTGLTT 708  
Qy 508 -----KYNLNGRESIINPGTA-----MASHKDEDEKFFPMG 538  
Db 709 CQKVDLDTKVINISIPKQINGSLNLTDNATANVKGLAKLNGVTLTNHSQ-----FTLS 763  
Qy 539 -----GVMIFGESAGASNTALD-NVMITDEEIKATN 570  
Db 764 NNATQIGNIRLSDNSTATVDNANLNGNVHLTDSAQFSLKN 803

RESULT 10  
US-09-815-242-5845  
; Sequence 5845, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; PRIOR FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5845  
; LENGTH: 1016  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-09-815-242-5845

Query Match 2.9%; Score 117; DB 10; Length 1016;  
Best Local Similarity 20.3%; Pred. No. 0.7;  
Matches 155; Conservative 96; Mismatches 272; Indels 240; Gaps 40;  
Qy 69 DAALEHDKAYDQOLKAGDNPYLRYNHADAEFOERLQEDTSFGGNLGRAVFOAKKRVL-- 126  
Db 1 EAAEAENNIENPTTLK--DN-----VQSKVEKIEEVTNKDTAPQGVAKSEVTSNKDTIEH 54  
Qy 127 EPLGLVEGAKTAPGKRPVEQSPQEPDSSSGIGKTGOQPAKKRL-----NFGQTGDS 180  
Db 55 EPSYKADISKEDTPKEVADVAEVQPKSS--VTHNAETPRVKARASVDEGSFDITRDSK 112  
Qy 181 SVPD-----PQPLGEPATPAVGPPTMASGGA PMAD----- 213  
Db 113 NVVESTPTITIGKEHFEQGYSDIQKPTDLGVSEVTRFNVGNESNGLIGALQKXKIDF 172

```
Qy 214 -----NNEGADG-----VGNASGNMHCDSITWLGDRVITSTRTWALPT 251
Db 173 SKDFNEKRVANNHQSNTTGGADGWGFLFSKGNAB-EYLTNGGILGDKGLVNS-GGFKIDT 230
Qy 252 ---YNNHLVKQISSASTGASNDNHFGYSTPWGYFDENRFHCHFSRDRWQRLINNNGWFR 308
Db 231 GYIYTSMDKTEKQAGQG-----YRGY-----GAF-----VKNDSSG-- 262
Qy 309 PKRLNFKLNIQVKEVTNDGVTITIANNLSTVQVFSDEYQ-----LPYVLGSAHQ 360
Db 263 ---NSQWGENIDKSKTN--FLNVAADNSTNT---SDGFHGORLNDVILTYVASTGKM 312
Qy 361 GC-----LPPFPADEVMIPOGYLTLNNGSQAVGRSSFYCYLXPSPQMLR 405
Db 313 RAAYAGKTWETSITDLGLSKNQAYNFLITSSQWGLNQGINANG-----WNR 359
Qy 406 T---GNNTFS-----YTFEEVPEHSSYAHQSLSLDRMLNPLIDOLYLYLNRTQN 451
Db 360 TDLKGSEFTTPPEAPKTTITELKKVEEIPFK-----ERKFNPDLPAGTEKVTR-BG 410
Qy 452 QSG-----SAQN-KDLLFSRGS-AGMSVQPKNMLP-----GPCYRQORVSKTKT 494
Db 411 QKGEKITTPTLKNPLTGVIISKGEPEEITKDPINELTEYGPETIAPGHRDEFDPKLP 470
Qy 495 DNNNSNFTWTGASKYNLNGRESIINP--GTAMASHKDEDEKFFPMSCGMIFGKESAGASN 552
Db 471 -----GEKEEVPKPGIKNPETGDIVRPPVDSVTKYGPVKGDSIVEKEE----- 514
Qy 553 TALDNVMTDEEIKATNPVA--TERFGTIVAVNFQSSST-----DPATGDVHAMGALPCM 605
Db 515 -----IPFEKERKFNPDLPAGTEKVTREGQKGEKITTPTLKNPLTGVIISKESKEE 567
Qy 606 VMQD--RDVYLQGPIMAKIPHDTGHPH-----SPLMGFGILKNPPQILIKNTPVPA 656
Db 568 ITKDPINELTEYGPETITPGHRD-EFDPKLPTEKEEVPKPGIKNPETGDIVR----- 620
Qy 657 NPPAESATKPFASFITQYSTGQVSEIEMELQKENSKRWNPEV 699
Db 621 -PPVD-SVTKYGPV-----KGDSIVEKE-EIPFEKERKFNPD 655

RESULT 11
US-09-815-242-12610
; Sequence 12610, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: BLITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
```

```
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12610
; LENGTH: 5795
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12610

Query Match      2.9%; Score 116.5; DB 10; length 5795;
Best Local Similarity 19.1%; Pred. No. 11;
Matches 128; Conservative 86; Mismatches 250; Indels 207; Gaps 35;

Qy 177 GDSSEVPDPQPLGEPATPAVGPPTMASGGAPMADNNEGADGVNASGNMHCDSITWLG 236
Db 1696 GNGSSAADYFKLSNGSAIPDA--TITWVGQAPNKNTRIGEDI-NVTANILID----G 1747
Qy 237 DRVITSTRTWALPTYNNHLVKQISSASTGASNDNHFGYSTPWGYFDENRFHCH----- 291
Db 1748 ETPITKTAT-----YKVS-----SVPKHVFETNRGAVFPGVSD 1782
Qy 292 -FSRDRWQRLINNNGWFRPKLNFKLNI-----QVKEVTNDGVTITIANNLSTVQVES 345
Db 1783 VYDAKQYVKPVNDSWTQNAQEMNFQFTNSYGPSKDVIGISTRDIRVTYDNHQTQIIKILA 1842
Qy 346 DSEYQLPYVLGSA-----HQCCLPFPADVEMIPQGYLTLNNGSQ 387
Db 1843 KYRDPDRIDGNSVTYKAGLTNQOIKINNVLSSSILFKAD--NTP----LTTNTTYG 1896
Qy 388 VGRSSFYCL-EYFSPQMLRTGNNFTFSYTFEEVPEHSSYAHQSLSLDRMLNPLIDOLYLYL 446
Db 1897 SGNATVAVTVSDALPNGVIKARS---SITMNNVTYTTQDEHGRAIDVTRNESVD----- 1946
Qy 447 NRTNQSQS-----AONKDLFSRGS-----PAGMSV---QPKW- 478
Db 1947 ---SNDSATVTVPOLQATTEGAVFIKGGDFDFGHVERFIQNPFGATVAMHDNPDTWK 2003
Qy 479 -----LPGPCYRQORVSKTKTDNNSNFTWTGASKYNLNGR 514
Db 2004 NTVGNTHTKTAVVTLPSSGGTRENVEVPVKVYPVANAKAPSRDVKQNLT-----NGT 2054
Qy 515 ESI-----INPGT-----AMASHKDEDEKFFPMSCGMIFGKESAGASNTALDNVMT-- 561
Db 2055 DAINVITFDPNTNTNGITAAWANNRQQPNQ---QAGVGHNVNDVTPGITAARVPTVN 2111
Qy 562 -DEEIKATNPVATERFGTIVAVNFQSSSTDPATGDVHAMGA--LP--GMV--WQDRDYYL 614
Db 2112 VYQFEFPQTSYTTTVG-GTLANGTQ-----ASGYAHMONANGLPTDFTYKMNNAATGT 2164
Qy 615 QGPIWAKI--PHT-----DGHFHPSPLMGGFGILKN-PPQILIKNTPVPA-- 656
Db 2165 NDANWAAAMKNPNAKVNAVYDVYNGHTFATSLPAKFVVKDVQPAKPTVETIAGAATI 2224
Qy 657 NPPAESATKPFASFITQYS-----TGQVSEIEMELQKENSKRWNPEVQYTSNYAKSA 709
Db 2225 TEGANQTVNTHAGNVTTYADKLVIKRGNV---VTIFTRRNTSPWKE---ASAATVA 2277
Qy 710 NVDFTVDNNGL 720
Db 2278 GIAGT--NNGI 2286

RESULT 12
US-10-022-809-2
; Sequence 2, Application US/10022809
; Publication No. US20030113407A1
; GENERAL INFORMATION:
; APPLICANT: LIN, Yi-Shin
; APPLICANT: LIU, Chang-Hsieh
; APPLICANT: CHU, Wen-Shen
; TITLE OF INVENTION: TRANSGLUAMINASE GENE OF STREPTOVERTICILLUM LADAKANUM AND THE
; TITLE OF INVENTION: TRANSGLUAMINASE ENCODED THEREFROM
; FILE REFERENCE: U-013779-2
; CURRENT APPLICATION NUMBER: US/10/022,809
```



CURRENT FILING DATE: 2002-03-29  
PRIOR APPLICATION NUMBER: US 10/021,678  
PRIOR FILING DATE: 2001-12-12  
NUMBER OF SEQ ID NOS: 3  
SEQ ID NO 2  
LENGTH: 410  
TYPE: PRT  
ORGANISM: Streptococcus lactis  
US-10-022-809-2

Query Match 2.9%; Score 115.5; DB 15; Length 410;  
Best Local Similarity 22.0%; Pred. No. 0.24;  
Matches 76; Conservative 43; Mismatches 151; Indels 75; Gaps 20;

QY 4 DGYLPDW-LEDNLSEGIRESW-WDLKPGAPKPKANQOKDDGRLVLPYKYLGPFG 58  
DB 97 DYPKPSYGRATITVNNYIRKQOVSHRDKRQOMTEBQREWLSTYGVVTWNSGQY- 154  
QY 59 LDKGEPVNAADAALHDKAYDQOLKAGDNPYLRYNHADAFAFQERLQEDT--SFGNLG 115  
DB 155 -----PTNRLAFPFDEDK-YKNELKNG-RP-RSGETRAFEGRVAKDSFDEAKGFQRA 205  
QY 116 RAVFOAKKRVLEPLGLVEEGAKTAPGKKR-----EVEQSPQEPDSS 156  
DB 206 RDVASVMNKALE--NAHDEGAYLDNLKELANGNDALRNEDARSPYSALRNTSPFKDRN 263  
QY 157 SGIGKTGQPAKKRL-----NFGOTGDSSEVPDPQLGEPRA--TPAVGPTTMASSGGA 209  
DB 264 GG-----NHDPKMKAVIYSKHFWSGQDRSGSDKRYGDPFAFRPDGGLVDMRDRI 319  
QY 210 PMADNEGADGVNAGSNMHCSTWLGDRTVITSTRT-WALPTYNNHLYKOISSASTGAS 268  
DB 320 PRSPSPGESFV-----NFDYGMFGAQTEADADKTVM---THGNHYH-----APNGSL 364  
QY 269 NDNH-YFGYSTPW--GYFDFNR--FHCHSPRDMQRL--INNHW 305  
DB 365 GAMHVESKFRNWSGDYSDFDGAYVVTFFPKSWNTAPDKVTQGW 409

RESULT 13  
US-10-170-682-3  
Sequence 3, Application US/10170682  
Publication No. US20030104975A1  
GENERAL INFORMATION:  
APPLICANT: Johan AUWERX  
APPLICANT: Pierre CHAMON  
APPLICANT: Frederic PICARD  
TITLE OF INVENTION: Cofactor-based screening method for nuclear receptor  
FILE REFERENCE: 017753-163  
CURRENT APPLICATION NUMBER: US/10/170,682  
CURRENT FILING DATE: 2003-01-02  
PRIOR APPLICATION NUMBER: US 60/297,772  
PRIOR FILING DATE: 2001-06-14  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: Patent Ver. 2.1  
SEQ ID NO 3  
LENGTH: 1441  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: SRC-1. Cloning of full-length human steroid receptor  
US-10-170-682-3

Query Match 2.8%; Score 112; DB 15; Length 1441;  
Best Local Similarity 19.3%; Pred. No. 3.2;  
Matches 140; Conservative 92; Mismatches 282; Indels 210; Gaps 32;

QY 48 PGYKYLGPFGNLDKGEFVNADAALHDKAYDQOLKAGDNPYLRYNHADAFAFQERL--Q 105  
DB 326 PSYRFI-----LNDGTMLSAHTKCKLCYQSPDMQ-----PFIMGIHIDREHSGLSPO 374

QY 106 EDTSEGNLGRAVFOAKKRVLEPLGLVEEGAKTAPGKKRVE-----QSPQEPD 154  
DB 375 DDTNSGMSIPR-VNPSVNPISPAHGVARSSTLPPSNMNSTRIHQSSDLHSSHN 433  
QY 155 SSGIGKTGQOP-----AKRLNFGQTDSESVDPQLGEPRAV-----GPTM 203  
DB 434 SSNSQGSFGCSPGSQIVANVALNKQASSQSSKPSLN-LNNPMEGTGISLAQFMSPRQ 492  
QY 204 ASGGGA--PMADNEGADGVNAGSNMHCSTWLGDRTVITSTRTWALPTYNNHLYKOIS 261  
DB 493 VTSGLATRPMPNNSFPNIST-----LSSPVGMTSSAC-----NNNRSYSNIP 537  
QY 262 SASGASND--NHFGYSTPMGFDFNRHCHFSPRDMQRLINNMGFRPKRLNFKLFI 319  
DB 538 VTSLOGMNEGPNNSVGFSA-----SPVLRQMSQNS-----PSRL-----NI 575  
QY 320 QVKEVTTNDGVTIANNTSTYQVFSSEYQLPYVLGSAHQCLPPFPADVFMIPQYGL 379  
DB 576 QPAKASKDN-KEIATSLNEMIQSDNSSSDGKPLDSGLH----- 614  
QY 380 TLNNGSOAVGRSSFYCLEYPSQMLRTGNFTFSYTFEEVPHSSYAHQSOLDRLNPLI 439  
DB 615 --NDRRLSDGSKYSQTSKILVQLLT-----TAEQQLRHA-----DIDTSCDVL 658  
QY 440 DQYLYLNRTONOSG-----SAQNKDL-LFSRGPAG--MSVOPKNWLPQCY 484  
DB 659 SCTGTSNSASANSNGSCSPSSHSLTARHKILHRLQEGSPSDITTLISVEP----- 709  
QY 485 RQQRVSKTDTNNSNFTWTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMGVM-- 542  
DB 710 -----DKDSASTSVSVTG-----QVQNSSI--KLELDASKKESKDHLRYLIDK 755  
QY 543 FGKESAGASNTALDNVMTDEEIKK-----TNPVATERFGTYAVNFQSSSTDPATGD-- 595  
DB 756 DEKDLSTPNLSLDVKKVKEKEQMDPCNTNPTPMTKPTPEIKLEAQSQTADLDQFD 815  
QY 596 ----VHMGALPGMWODRDVYLOGPIWAKIPHTDGHFHSPLMGFGGLKNPPQILIK 650  
DB 816 QLPTEKAAQLPGLCETDR--MDGAV-----TSVTIK 846  
QY 651 NTPVPANPPAEFSATKFASTFOYSTGQSVIEIWELOKNSKRNPP--EVQYTSNYA 706  
DB 847 SEILPAS-----LQSAIAPRTSRLNRLPELELAIDNQFGQGTGDQIPWNTV 896  
QY 707 KSAAN 710  
DB 897 TAIN 900

RESULT 14  
US-09-862-027-80  
Sequence 80, Application US/09862027  
Patent No. US20020142428A1  
GENERAL INFORMATION:  
APPLICANT: Hodge, Martin R.  
TITLE OF INVENTION: No. US20020142428A1 Kinases and Uses Thereof  
FILE REFERENCE: 35800/234862  
CURRENT APPLICATION NUMBER: US/09/862,027  
CURRENT FILING DATE: 2001-05-21  
PRIOR APPLICATION NUMBER: US 09/345,473  
PRIOR FILING DATE: 1999-06-30  
NUMBER OF SEQ ID NOS: 82  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 80  
LENGTH: 1265  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-862-027-80

Query Match 2.8%; Score 111.5; DB 11; Length 1265;  
Best Local Similarity 24.9%; Pred. No. 2.9;  
Matches 57; Conservative 27; Mismatches 70; Indels 75; Gaps 12;

Search completed: July 17, 2003, 18:51:13  
Job time : 39.4414 secs

QY 56 ENGLD-----KGPVNADAAALHDKAYDQOLKAGDNPYLRYNHADAEEFQERLOED- 107  
DB 316 FTGLDWTGLRQKAEFTIPOLSE--EDDTSY-----FDRSERHYHMDSEDEEVSSEDG 366  
QY 108 ----TSFGNLGRAVFAQAKRVLEPLGLVEGAKTAPGKRPVEGSPQEPDSSSGI-GKT 162  
DB 367 CLEIRQFSSCSPPR--FNKVYSMERLSTLEE-RTTPPTKRSL--SEKEDHSDGLAGLK 421  
QY 163 GQQPA-----KKRLNFQGTGDSSEVPDP----- 185  
DB 422 GRDRSWVIGSPETLRKRLSVSESSHTESDSSPMTVRRRCGGLDAPRPEGPBEASSTL 481  
QY 186 --QP-----LGEPPATPAVGPPTMASGGGAPMADNNEGADGVGNASGN 227  
DB 482 RRQPOEGIWLTLPSPGEGVSGPVTEHSGEQRPKLDE---EAVGRSSGS 526

RESULT 15  
US-09-859-888-4

; Sequence 4, Application US/09859888  
; Patent No. US20020173459A1  
; GENERAL INFORMATION:  
; APPLICANT: YAN, Chunhua et al  
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND  
; TITLE OF INVENTION: USES THEREOF  
; FILE REFERENCE: CL001239  
; CURRENT APPLICATION NUMBER: US/09/859,888  
; CURRENT FILING DATE: 2001-05-18  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 433  
; TYPE: PRT  
; ORGANISM: Human  
US-09-859-888-4

Query Match 2.8%; Score 111; DB 11; Length 433;  
Best Local Similarity 23.1%; Pred. No. 0.63;  
Matches 101; Conservative 40; Mismatches 125; Indels 172; Gaps 27;

QY 77 KAYDQOLKAGDNPYLRYNHADAEEFQERLOEDTSFGNLGRAVFAQAKRVLE--PLGLVE 133  
DB 80 RSLDAQLRAD-----GSLSAKSFQELKDRMELLPLSSVL 115  
QY 134 EGAKTAPGKRPVEGSPQEPDSSSG-----IGKTGQQPAKKR-----L 171  
DB 116 EYK---ADTRTIVRLREVERNLGSLAALQEMGAYGEDLQQRVWALEARLHACAQKL 172  
QY 172 NFGQ--TGDSSEVPDPQPLGEPATPAVG-----PTMASGGGAPMADNNEGADGVGN 223  
DB 173 GCGKLTGVS-----PITVRAMGSRFGSWMTDTM-----APRAD----- 206  
QY 224 ASGNWHCDSTWLDREVITTTSTRTWALPTVN--NHLVKQISSASTGASNDNHVFGYSTPW- 280  
DB 207 -SRVWYMDGYKGRVLEFRTLGDPIKQNFIOHLLPQ-----PWA 246  
QY 281 -----GYFDNRPHC-----HFSPRD--WQRLI-----NN-----NWGFRPKRLNFK 315  
DB 247 GTGHVVYNGSLFYNKYQSNVVKHFRSRVYLVQKSLPGAGYNNTPPYSWG-----GFS 300  
QY 316 LENIQVKE-----VTNDGVTTIANNLSTGVQVFSDEYQLPYVLGSAHQGCLPPFPAD 369  
DB 301 DMDFMVDESGLMAVYTTNQN---AGNI-----VVSRLDPHTLEWRSWDTGYPKRSAGE 351  
QY 370 VFMIPOGYLTLNNGSQAVGRSSFYCLBEPFSQMLRTGNNFTFSYTFEEVFPFHSSYASQ 429  
DB 352 AFMICGVLVYT---NSHLAGAKVYFA--YFTN-----TSSYEYTDVPEFHNOYSHIS 397  
QY 430 SLDRLMNPLIDQVLYLIN 447  
DB 398 MLD--YNPR-ERALYTNW 412





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 2003, 18:28:05 ; Search time 26.3842 Seconds\*  
(without alignments)  
2681.720 Million cell updates/sec

Title: US-09-807-802A-13

Perfect score: 3989

Sequence: 1 MADGXLPDWLEDNLSEGIR.....NNGLYTEPRPIGTRYLTRPL 736

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 73:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2177.5	54.6	504	1	VCPV3A coat protein - ade
2	2168.5	54.4	732	2	SS2210 coat protein VP1 -
3	798.5	20.0	673	1	VCPVB5 coat protein VP1 -
4	600.5	15.1	781	1	VCPV19 coat protein VP1 -
5	482	12.1	729	1	VCPVNA coat protein VP1 -
6	482	12.1	729	1	A60006 coat protein VP1 -
7	444.5	11.1	727	1	VCPV1F coat protein VP1 -
8	439.5	11.0	722	1	VCPVME coat protein VP1 -
9	431.5	10.8	727	1	VCPVFP coat protein VP1 -
10	426	10.7	723	1	VCPVPP coat protein VP1 -
11	424.5	10.6	718	1	VCPVIM coat protein VP1 -
12	421.5	10.6	748	1	VCPVCP coat protein VP1 -
13	407.5	10.2	737	1	VCPVCD coat protein VP1 -
14	404	10.1	722	1	VCPVV2 coat protein VP1 -
15	385.5	9.7	722	1	VCPVCN coat protein VP1 -
16	377	9.5	716	1	VCPV2M coat protein VP1 -
17	234	5.9	587	1	B44276 coat protein VP1 -
18	199.5	5.0	810	2	A44054 coat protein VP1 -
19	198	5.0	584	2	S49594 coat protein VP1 -
20	191	4.8	702	1	VCPVAP coat protein VP1 -
21	143.5	3.6	648	2	S50856 whn protein - rat
22	139	3.5	931	2	T49710 related to glucan
23	131.5	3.3	1446	1	A45344 immediate-early pr
24	130.5	3.3	1145	2	T18235 transcription acti
25	130	3.3	1338	2	T30565 MAP kinase kinase
26	129	3.2	1072	2	A86827 hypothetical prote
27	124.5	3.1	2493	2	A55481 adenylate cyclase
28	124	3.1	1296	2	C82521 hemolysin-type cal
29	124	3.1	3078	2	T28432 variant-specific s

30	123.5	3.1	690	2	AB0124 probable TonB-depe
31	123.5	3.1	2529	2	B64635 toxin-like outer m
32	123	3.1	1142	2	T37455 enamelrin precursor
33	122.5	3.1	880	1	SYBSVS valine-tRNA ligase
34	121	3.0	635	2	F96660 protein F2K1.10 f
35	121	3.0	1113	2	S28925 nuclear pore compl
36	121	3.0	1549	2	I48103 type VII collagen
37	120	3.0	1379	2	T37752 hypothetical serin
38	119.5	3.0	1180	2	E86719 hypothetical prote
39	119.5	3.0	1394	2	S60762 Iga-specific serin
40	119.5	3.0	2004	2	AC0314 probable membrane
41	119	3.0	667	2	A41311 transcription fact
42	118.5	3.0	2399	2	H71879 toxin-like outer m
43	118	3.0	739	2	T52289 probable transkeo
44	118	3.0	956	2	T08144 myrosinase-binding
45	117.5	2.9	1076	1	A35622 nuclear pore prote

ALIGNMENTS

RESULT 1									
VCPV3A									
coat protein - adeno-associated virus type 2									
C:Species: adeno-associated virus type 2									
C:Date: 05-Apr-1983 #sequence_revision 05-Apr-1983 #text_change 16-Jul-1999									
C:Accession: A03698									
R:Srivastava, A.; Lusby, E.W.; Berns, K.I.									
J. Virol. 45, 555-564, 1983									
A:Title: Nucleotide sequence and organization of the adeno-associated virus 2 genome									
A:Reference number: A03694; MUID:83164299; PMID:6300419									
A:Accession: A03698									
A:Status: translation not shown									
A:Molecule type: DNA									
A:Residues: 1-504 <SRI>									
A:Cross-references: EMBL:J01901; NID:g209616; PIDN:AAA42376.1; PID:g209621; EMBL:M12405;									
C:Superfamily: adeno-associated virus coat protein									
C:Keywords: coat protein									
Query Match									
Best Local Similarity 54.6%; Score 2177.5; DB 1; Length 504;									
Matches 392; Conservative 33; Mismatches 47; Indels 3; Gaps 2;									
QY	203	MASGGAPMADNNEGADGVGNASGNWCHDSTWLGDRTTSTRTWALPTYNNHLYKQISS	262						
DB	1	MATSGGAPMADNNEGADGVGNSSGNWCHDSTWMDRVTITSTRTWALPTYNNHLYKQISS	60						
QY	263	ASTGASNDNHFGYSTPWGVEDFNRFCHFSPRDWORLNNWGFRRPKRLNFKLFNIQVK	322						
DB	61	QS-GASNDNHFGYSTPWGVEDFNRFCHFSPRDWORLNNWGFRRPKRLNFKLFNIQVK	119						
QY	323	EVTNDGVTTIANNLSTVQVFSDSXYQLPYVLGSAHQGLPPFPADVFMIPQYGLTLN	382						
DB	120	EVTQNDGTTTIANNLSTVQVFTDSEYQLPYVLGSAHQGLPPFPADVFMIPQYGLTLN	179						
QY	383	NGSQAVGRSSFYCLEYFPSSQMLRTGNNFTFSYTFEEVFPFHSSYAHSQSLDRLMPLIDQY	442						
DB	180	NGSQAVGRSSFYCLEYFPSSQMLRTGNNFTFSYTFEEDVFPFHSSYAHSQSLDRLMPLIDQY	239						
QY	443	LYYINRTQNSGSAQNKDLFSRGSPPAGMSVQPKNWLPGPCYRQORVSKTKTDNNSNFT	502						
DB	240	LYYLSRTNTPSGTTTQSRLOFSQAGASDIRDQSRNWLPGPCYRQORVSKTSADNNSNFT	299						
QY	503	WTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMGSMVIFGKESAGASNTALDNVMTD	562						
DB	300	WTGATKYHLNGRDSLVP--AMASHKDEDEKFFPQSGVLIIFGKQSEKTNVIEKVMITD	357						
QY	563	EEETKATNPVATERFGTVAVNFOSSSTDTPATGVDHAMGALPGMWQDRDVILOGPIMAKI	622						
DB	358	EEETGTTNPVATEQYGSVSTNLQGRNQAATADVNTQGVLPGMVWQDRDVILOGPIMAKI	417						
QY	623	PHTDGHFHPSPLMGGFGLKNDPPQILIKNTVPANPPAFSATKPFASFITQYSTG	677						

Db 418 PHTDGHFHPSPLMGGFGLKHPPQILLKNTPVDPANPSTTFSAKFAFSFITQYSTG 472

RESULT 2

S52210

coat protein VP1 - muscovy duck parvovirus

N;Alternate names: VP1 protein

C;Species: muscovy duck parvovirus

C;Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 17-Mar-2000

C;Accession: S52210

R;Zadortl, Z.; Erdei, J.; Nagy, J.; Kisari, J.  
submitted to the EMBL Data Library, September 1993

A;Reference number: S52209

A;Accession: S52210

A;Molecule type: DNA

A;Residues: 1-732 <ZAD>

A;Cross-references: EMBL:X75093; NID:g609091; PIDN:CAA52384.1; PID:g609093

A;Experimental source: strain FM

C;Genetics:

A;Gene: VP1

C;Superfamily: parvovirus coat protein

C;Keywords: coat protein

Query Match	54.4%;	Score 2168.5;	DB 2;	Length 732;	.
Best Local Similarity	55.2%;	Pred. No. 8.1e-134;			
Matches 418;	Conservative 90;	Mismatches 186;	Indels 63;	Gaps 13;	

```

Qy      9 DWLEDNLSEGI REMWDLKPGAPKPKANQOKQ-----DDGRGLVLPGYKYLGPEN 57
        |||      ||| |||||::|||      ||| |||||::|||
Db     10 DWYE-----TAASWRHLKAGAPKPKSNQOSQSVSTDRKPQRKDNNGRGVFLPGYKYVGPGN 65

```

```

Qy      58 GLDKGEPVNADAAL EHKDAYDQOLKAGDNPYLRYNHADAEFOERLOEDTSFGGNLGRA 117
      ||||| : ||||| ||||| ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db      66 GLDKGPPVNKADSVALEHKDAYDQOLKAGDNPYIKFKHADQEFITNLQGDTSFGGNL GKA 125

```

```

Qy      118 VFQAKRVLLEPLGLVEEGAATA PGKKR PVEQS PPEP DSSSG IGTGQ QPAKKRLNFGQTG 177
        |||||:|||||||  |||  |||  |  |  :
Db      126 VFQAKRRILEPLGLVEEPPVNTAPAKK-----SSG-KLT DHDPIVKKPKLSE-- 170

```

```

QY      178 DSESVPDPQPLG-----EPATPAVGPPTTMASGGCAPMADNNEGADGVGNASG 226
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      171 --ENSPSPNSGGEASAAATEGSEPYAP-----NMAEGGSGAMGDSAGGADGVGNASG 222

```

```

QY      227  NWHCDSTWLGDRVITTTSTRTWALPTNNHNLKQISSASTGASNDNHFGYSTPMGYFDEN 286
        ||||| ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||
Db      223  NWHCDSTWLGDRVITTKTRTWALPSYNNHMYOAITSGTNPDSN -TOYAGYSTPMGYFDEN 281

```

```
QY      287 RFHCHFSPRDWÖRLINNNGWFRPKRLNFKLFNIÖVKEVTNIGVTIANNLSTVÖVFS 346
        |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      282 RFHCHFSPRDWÖRLINNHWGIRPALKEKIFENVÖVKEVTTOPÖRTIANNLSTIOIFTD 341
```

```
QY      347 SEYQLPYVLGSAHQGLPPFPADVFMIPQGYLTLN--NGSQAVGRSSFYCLEYPSPQM 403
        ::|||::| : ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      342 NEHQLPYVIGSATEGTMPPEPSPDYALPOYGCTMHTNOSGARENDRSAFYCLEYFSPQM 401
```

OY 404 LRTGNFTSFYTFFEEVPPHSSVAHSQSLDRIMNPLIDQYLVLNPTONOGSQAOKDLLE 463  
| | | | : | | | | : | | | | | | | : | |  
Db 402 LRTGNNPFPSPFPEEFVPFHSMFAHSODLIRIMNPILDOYLMIFFSEV-NGGRNAO-----F 455

OY 464 SRGSPAGMSVOPKMWLPBPCRYQORYSKTk--TDNNSNSFTWTGASKYNLNGRESIINPG 521  
:  
456 :  
KRKVKGACGAMGPNNMTDGGPKLIDOBVPAYSGCTDNVYANWSTWSKGNKYETKDREVTILOPG 515

[illegible]

Qy 580 VAVNFQSSSTDPATGDIVHAMGALPGWVWQDRDVYLQGPITWAKLPHTDGHHFHPSPMLMGFFG 639

QY 640 LKNPPQILIKNTPVVPANPPAESATKFAFITQYSTGQVSVEIEMELQKENSKRWNPEV 699  
| | | | | : | | | | | : | : | : | | | | | : | | : | | | | | | | | :

Db | 636 LHNPPQYFIKNTFVPADPPLLEYNAQKNSYITQYSTGQCYEMVWELRKENSKRWNPET 6935

QY | 700 QYTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 736

Db | 696 QFTSNFGNSTMFAPNETGGYVEDRLIGTRYLTQNL 732

RESULT 3  
VCPIVB5

coat protein VP1 - bovine parvovirus  
N/Contains: coat protein VP2  
C/Species: bovine parvovirus  
C/Date: 28-Dec-1987 #sequence\_revision 28-Dec-1987 #text\_change 16-Jul-1999  
C/Accession: A26104  
J/Chen, K.C.; Shull, B.C.; Moses, E.A.; Lederman, M.; Stout, E.R.; Bates, R.C.  
J./Virol. 60, 1085-1097, 1986  
A/Title: Complete nucleotide sequence and genome organization of bovine parvovirus  
A/Reference number: A26104; MUID:87061184; PMID:3783814  
A/Accession: A26104  
A/Molecule type: DNA  
A/Residues: 1-673 <CHE>  
A/Cross-references: EMBL:M14363; NID:g333454; PIDN:AAB59847.1; PID:g808805  
C/Superfamily: parvovirus coat protein  
C/Keywords: coat protein  
F;138-673/Product: coat protein VP2 #status predicted <VP2>

Query Match	20.0%;	Score 798.5;	DB 1;	Length 673;
Best Local Similarity	28.6%;	Pred. No. 2.4e-44;		
Matches 214;	Conservative 105;	Mismatches 275;	Indels 155;	Gaps 21;

Qy	Db	30	3	57
PKPKANQOKDDEGRGLVLPGYKYLGPFGNLGKGEPEYNAADAAL	GLTLPGYNYLGPFNSLFAGAPVNVKADAAARKHDFGYSDDLKEGKNP	EHDKAYDQQLKAGDNP		

```
QY      90 YLRYNHADAEFQERLQEDTSTFGGNLGRAVPQAKKRVLEPLGLVEGAKTAPGKKRPVEQS   149
       ||::||| : ::||||| | | | | | : : : :
Db      58 YLYENTHDONLIBELKDDTSTFGGKLARGVFOIKKALAPAL-----PGTSKGDRALKRK   111
```

QY	150	PQEPDSSSGIGKTGQQPAKRRLNFGQYGDSESVDPQPLGEP--PATPAAVGPTTMASGG	207
		:     : :     : : :     :     :	
Db	112	LYPARSNKGAKKANREBPASTSNQOMEVSNIDIPDEAGNQPIELATRSVVGSGSVGGG	171

QY		208 GAPMADNNEGADVGNASGNMHCDSTWLGDRTVTTSTRWALPTYNHHLKYQISSASTGA	267
		::  :::  ::   :	:
Db	.	172 -----RGGSGVGYSTGWTGTTFSENIVATKNTROFICDIKNGLYKS-EVLNTGD	222

QY	268	SNDNHYFGYSTPWCYEDFNRFHCHFS	PRDWO	RLNNNNWGFRPKRLNFKL	FNIOQKEV	TTN	327
	:	:	:	:	:	:	:
Db	223	TAHROY-AITTPWSYENENYOSSHS	FSPNDWOHLV	YERFRPKAMIV	RVVNLQIKO	IMTD	281

QY	382	-NNGSQA VGRS-----	SFYCLEYFP SQMLRTGNFTFSYTFEEVFPFHSSYAH SQSLDR LM	435
		:	:	:
Dh	342	DNNTTNTVFEETLTGVP L YML ENSDHEV L RNG-----	-----R IY	376

436	NPLIDQYLTYLNRTON-----QSGSAQNKKLLFSRGSP-----AGMSVQPKMMLP	480
437	PIYTOIWMPIRMDEKONHTIOHSDNDVOSTGOKOKNTLTOTKRONKORFONAAIPTSNNMS	436

QY	481	GPCYRQQRVSKTYTDNNNSNFTWTGASKYNLNGRESTINPCTAMASHKDEDKFFPMSCV	540
		:: : :	
DB	437	CD--GAPCTUNNTVOTOSGAGIYMMATNGAD--YSGV	470

QY	541 M-----IFGKE-----SAGASNTALDNVMTDDEEIKATNPVATE	575
	:   :	
05	421 DATTUVCCTMPDTCCOONEDTIDPDKCAKAECONNITEN	516

576 REGTAVVNFGSSSTDPA TGDVHAMGALPGMWQDRVYLLGP I WAKI PHTDGHFHPSP LM 635

QY 636 GGFGKPNPPQILIK--NTPVPANPPAEFSATKFASTITQYSTGQSVIEIEMELQKENS 693  
 Db 577 GSIPMSHPGTFITKLARIIPVPGNGD-----SFLNITYVTGQSVCEVWEVEKRGTK 627  
 QY 694 RMNPEVOYTSNVAKSANVD-FTVDNNGLY 721  
 Db 628 NMRPEYMHS--ATNMSVDAYTINNAGVY 653

RESULT 4

VCPV19  
 coat protein VP1 - parvovirus B19 (strain Au)  
 C:Species: parvovirus B19  
 A>Note: host Homo sapiens (man)  
 C>Date: 28-Dec-1987 #sequence\_revision 28-Dec-1987 #text\_change 16-Jul-1999  
 C:Accession: A24299  
 R:Shade, R.O.; Blundell, M.C.; Cotmore, S.F.; Tattersall, P.; Astell, C.R.  
 J. Virol. 58, 921-936, 1986  
 A>Title: Nucleotide sequence and genome organization of human parvovirus B19 isolated fr  
 A:Reference number: A24299; MUID:86200451; PMID:3701931  
 A:Accession: A24299  
 A:Molecule type: DNA  
 A:Residues: 1-781 <SHA>  
 A:Cross-references: EMBL:ML3178; NID:g333375; PIDN:AAA66867.1; PID:g333377  
 C:Superfamily: parvovirus coat protein  
 C:Keywords: coat protein

Query Match 15.1%; Score 600.5; DB 1; Length 781;  
 Best Local Similarity 26.6%; Pred. No. 2.6e-31;  
 Matches 182; Conservative 95; Mismatches 300; Indels 107; Gaps 20;

QY 47 LPGYKYLGPFGNLKGEFVNADAAALHDKAYDQQLKAGDNPYLRYNHADAFFQERLOE 106  
 Db 125 LPGTNYVGPFGNELQAGPPQSAVDSAAARIHDFRYSQQLAKGINPYTHWTVADEELKNIKN 184  
 QY 107 DTSFGNLGRAVFOAKKRVLEPLGLVEGAKTAPGKKRPVEQSPQEPDSSSGIGKTGQOP 166  
 Db 185 ETGFOAQVNVKDYF-----TLKGAAPVAHFQ-----GSLPEVP 217  
 QY 167 AKKRLNFGQTSSESVPDQPLGEPATPAVGPPTMASGGAPMADNNEGADGVGNASG 226  
 Db 218 AYNA-----SEKYPMSITSVNSAEASTGA-----GGG-----GSNSVKS 250  
 QY 227 NMHCDSTWLGDVITTTSTRTWALPTYNHNLKQISSASTGASND-----NHYPGY 276  
 Db 251 MMSEGATFSANSVTCTFSRQFLIPYDPEHHYKVFSPAASSCHNASGKAACVCTISPIMGY 310  
 QY 277 STPWGYFDENRHHCHSPRDWQRLINNNWGRPRKLNFKLFNIQVKEVT--TNDGVTTIA 334  
 Db 311 STPWRYLDFNALNLFSPLEFQHLIENYGSIAADALTVTISEIAVNDVTDKGGGV-QVT 369  
 QY 335 NMLSTVQVFSSEYQLPYVLGSAHQGCLPEPADVFMIPQGYLTINN-GSQAVG---- 389  
 Db 370 DSTGRLCMVDHEKYPYVLGGQDITLAPLPIWVYFPQYAYLTWGDVNTQGISGDSK 429  
 QY 390 -----RSSFCLEYFSPQMLRTGNNFTFSYTEEVPFHSYAHQSGLDRMLNPLIDQYLY 444  
 Db 430 KLASSESAFYVLEHSSFQLLGTGTASMSYKPPVPENLEGCSQHFYEMYNPL---YGS 486  
 QY 445 YLNRTONQSGSAQNKDILFSRGSAPGMSVQPKWLPGPCYRQQRVSKTKTDNNNSNFTWT 504  
 Db 487 RLGVDPDTLGGDPKFRSL-----THEDHAIDQONFMPGPLVNSVSTKGDSSNTGAKALT 541  
 QY 505 GASKYNLNGRESLIINPG-TAMASHKDEDEKFFPMGCVMIFGKESAGASNTALDNV----- 558  
 Db 542 GLSTGTSQNTIRISLRFPVSQPYHHWDTDKYVTGINAISHGQTTYG---NAEDKEYQGGV 598  
 QY 559 --MITDEEIKATNPVATERFGTVAVNFQSSSTDPATGVDHAMGALPGWVQDRDVTYLOG 616  
 Db 599 GRFPNEKEQLKQLOGINMHTY-----FPNKGTQQYTDQIE-RPLMVGSVWNRRLAHYES 651  
 QY 617 PIWAKIPIHTDGHFHS-PLMGFGKLNPPQILIKNTVPANPPAEFSATKFASTITQYS 675

Db 652 QLMKSKINLDDSFKTQFALGGWGLHQPPQIFLK--ILPQSGPIGIGIKSMGITTLVQYA 709  
 QY 676 TGQVSVIEIEMEL-QKENSKRWNPE 698  
 Db 710 VGIMVTWMTFKLGPRAKATGRWNPQ 733

RESULT 5

VCPVNA  
 coat protein VP1 - porcine parvovirus (strain MADL-2)  
 N:Contains: coat protein VP2  
 C:Species: porcine parvovirus  
 C>Date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 16-Jul-1999  
 C:Accession: B33743; D48472  
 R:Vaudevacharya, J.; Basak, S.; Srinivas, R.V.; Compan, R.W.  
 Virology 173, 368-377, 1989  
 A>Title: Nucleotide sequence analysis of the capsid genes and the right-hand terminal pa  
 A:Reference number: A33743; MUID:90085785; PMID:2596019  
 A:Accession: B33743  
 A:Molecule type: DNA  
 A:Residues: 1-729 <VAS>  
 A:Cross-references: GB:M32787; NID:g332983; PIDN:AAA46917.1; PID:g332985  
 R:Bergerson, J.; Meneses, J.; Tijssen, P.  
 Virology 197, 86-98, 1993  
 A>Title: Genomic organization and mapping of transcription and translation products of t  
 A:Reference number: A48472; MUID:94025614; PMID:8212598  
 A:Accession: D48472  
 A:Molecule type: DNA  
 A:Residues: 11-729 <BER>  
 A:Experimental source: NADL-2, ATCC VR-742  
 A>Note: sequence extracted from NCBI backbone (NCBIN:138789, NCBI:138794)  
 C:Genetics:  
 A:Introns: 10/1  
 C:Superfamily: parvovirus coat protein  
 C:Keywords: coat protein; glycoprotein  
 F:151-729/Product: coat protein VP2 #status predicted <VP2>  
 F:32,172,198,282,330,433,471,573,604,651/Binding site: carbohydrate (Asn) (covalent) #ste

Query Match 12.1%; Score 482; DB 1; Length 729;  
 Best Local Similarity 25.5%; Pred. No. 1.3e-23;  
 Matches 197; Conservative 111; Mismatches 310; Indels 154; Gaps 35;

QY 43 RGLVLPYKYLGPFGNLKGEFVNADAAALHDKAYDQQLKAGDNPYLRYNHADAFFQ 102  
 Db 9 RGLTLPYKYLGPFGNSLDQGEPTNPSDAAKAHEADYDKYIKSGKNPYFYFSAADEKFIK 68  
 QY 103 RLQEDTSFGNLGRAVFOAKKRVLEPLGLVEGAKTAPGKKRPVEQSPQEPDSSSGIGK- 161  
 Db 69 ETEHAKDYGKIGHYFFRAKRAFAFKL---SETDSPTTSQQPEVRKSPKHPGSKPPGKR 125  
 QY 162 -----TGQOPAKRLNFGQTSSESVPDQPLGEPATPAVGPPTMASGGGAPMA 212  
 Db 126 PAPRHIFINLAKKAKAGTSNTNSMSSENVQHNPIINAGTEL-SATGNESGGGGGG--- 181  
 QY 213 DNNEGADGVGNASG--NMHCDSTWLG--RVITSTRTWALPTYNHNLKQIS--SAST 265  
 Db 182 -GGRGAGGVSTGTFFNQTEFQYLGEGLVRITAHASRLIHLNMPHEHETKYRIHVLNSES 240  
 QY 266 GAS-----NDNHYFGYSTPWGYFENRHHCHSPRDWQRLINNNWGRPRKLNFKLFNIQV 321  
 Db 241 GVAQGVQDDAHTQMTVPMSLIDANAMGWENPADWQLISNNMTEINLVSEQEIFNVVL 300  
 QY 322 KEVT--TNDGVTTIANNLSTVQVFSSEYQLPYVLGSAHQGCLPEPADVFMIPQGY 378  
 Db 301 KITESATSPPTKIYNNDLFTASLWALDNTNLTLPYTPAARSETLGFYPLPTKPTQYRY 360  
 QY 379 L-----TLNNGSOAVGRS-----SFYCLE-YFPSQMLRTGNNF-TFSYTE 417  
 Db 361 YLSCIRNLNLPPTYTGSSQQTIPSIOGTGLHSDIMFYTIENAVPIHLIRTGDEFSTGIYHD 420  
 QY 418 EVPFHSSYAHQSGLDR-----LMNPILI--DQYLYLNRTONQSGSAQNKDILFSRGS 467



Db 421 TKPL--KLTHSWQTNRSGLPRLKLTPTTEGDQHPGTLPAANTRKGYHOTINNSYTEAT 478  
Qy 468 PAGMSVQPKNWLPGPCYRQQRVSKTKTDNNNSNFTWTGASKYNLNGRESIINP--GTAMA 525  
Db 479 ----AIRP-----AQVGYNTPYMNFYSNGGPF-----LTPIVPTADT 512  
Qy 526 SHKDEDEKFFPMGSGVMIFGKESAGASNTALD---NMWITDEEIK--ATNP-----VAT 574  
Db 513 QYNDEPN-----GAIRFTMDYQHGHLTTSQELERYTFNPQSKGRAPK 557  
Qy 575 ERFGTVA-VNFOSSS-----TDPATG--DVHAMGALP-----GMWQDRDYYLQGPIMA 620  
Db 558 QQFNQAPLNLENTNNGTLLPSDPIGGKSNMFMNTLNTYGPLTALNNTAPVFPNGQIMD 617  
Qy 621 KIPHTD--GHFHPSPLMGGFGLK-NPPQILIKNTPPVPANPFAEFS--TKFASFITQYST 676  
Db 618 KELDTDLKPRLH---VTAPFVCNNPFGQLFVKIAP---NLTDENADSPOQPRIITYSN 671  
Qy 677 GQVSVEIEWELOKENSKRWNPEVOYTSNYAKSANVDFTVDNGLYTEPRPIG 728  
Db 672 FMMKGTLTFTAKRSSNMWNP IQOHTT-----TAENIGNYI-PTNIG 712

RESULT 6  
A60006  
coat protein VP1 - porcine parvovirus (strain 90HS)  
N;Contains: coat protein VP2  
C;Species: porcine parvovirus  
C;Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 30-Sep-1993  
C;Accession: A60006  
R;Sakurai, M.; Nishimori, T.; Ushimi, C.; Nakajima, H.  
Virus Res. 13, 79-86, 1989  
A;Title: Nucleotide sequence of capsid protein gene of porcine parvovirus.  
A;Reference number: A60006; MUID:89319168; PMID:2750278  
A;Accession: A60006  
A;Molecule type: DNA  
A;Residues: 1-729 <SAK>  
C;Superfamily: parvovirus coat protein  
C;Keywords: coat protein; glycoprotein  
F;151-729/Product: coat protein VP2 #status predicted <VP2>  
F;172,198,282,330,433,471,573,604,651/Binding site: carbohydrate (Asn) (covalent) #statu

Query Match 12.1%; Score 482; DB 1; Length 729;  
Best Local Similarity 25.7%; Pred. No. 1.3e-23;  
Matches 197; Conservative 112; Mismatches 314; Indels 144; Gaps 35;

Qy 43 RGLVLPGYKYLGPENGLDKGEPVNAADAALHDKAYDQOLKAGDNPYLRYNHADAEPQE 102  
Db 9 RGLTLPGYKYLGPENSLDGEPTNPSDAAKEHDEAYDKYIKSGKNPYFYSAADEKFIK 68  
Qy 103 RLQEDTSFGNLRGAVFOAKKRVLEPLGLVEBGAKTAPGKRPEVQSPQEPDSSSGIGK- 161  
Db 69 ETEHAKDYGKIGHYFFRA-KRAFRP-KLSETDSPTTSQOPEVRRSPRKHGSKPPGKR 125  
Qy 162 -----TGQPAKKRLNFGQTGDSSEVPDQPLGEPPTPAVGPPTMASGGGAPMA 212  
Db 126 PAPRHIFINLAKKAKGTSNTNSMSSENVQHNPIINAGTEL-SATGNESGGGGGG-- 181  
Qy 213 DNNEGADGVNAGNWH--CDSTWLGD---RVITSTRTWALPTYNNHLYKOIS--SAST 265  
Db 182 -GGRGAGGVSTGSFNNQTBFOYLGEGLVRIATAHASRLIHLNMPHEHYKRIHVLNSES 240  
Qy 266 GAS----NDNHYFGYSTPWGYFDENRFHCHFSPRDMQRLINNMWGFPRKRLNFKLENIOV 321  
Db 241 GVAGQMVDDAHITQWTPWLSLIDANAMGWENPADWQLISNMTEINLVSEQEIFNVVL 300  
Qy 322 KEVT---TNDGVTIANNLSTVQVFSDEYQLPYVLGSAHQGLPPFPADVEMIPQYGY 378  
Db 301 KTTTESATSPPTKIYNNDLTASLVALDTNNTLPTTPAAPR3ETLGFYMWLPTKPTQYRY 360  
Qy 379 L-----TLNNGSQAVGRS-----SFYCLE-YFPSQMLRTIGNNF-TFSYTFE 417  
Db 361 YLSCTRLNLPPTYTGOSQOITDSIQTGLHSDIMFYTIENAV?IHLLRRTGDEFSTGIYHFD 420

Qy 418 EVPFHSSYAHSQSLDR-----LMNP LI--DQYLYLNRITONQSGSAQNKDLFESRGS 467  
Db 421 TKPL--KLTHSWQTNRSGLPRLKLTPTTEGDQHPGTLPAANTRKGYHOTMNSYTEAT 478  
Qy 468 PAGMSVQPKNWLPGPCYRQQRVSKTKTDNNNSNFTWTGASKYNLNGRESIINP--GTAMA 525  
Db 479 ----AIRP-----AQVGYNTPYMNFYSNGGPF-----LTPIVPTADT 512  
Qy 526 SHKDEDEKFFPMGSGVMIFGKESAGASNTALDNDNMWITDEEIKATNP-----VATERFGT 579  
Db 513 QYNDE-----PNGAIRFTMGYQHQLTTS-----SQELERYTFNPQSKGRAPKQFNQ 562  
Qy 580 VA-VNFOSSS-----TDPATG--DVHAMGALP-----GMWQDRDYYLQGPIMAKIPHT 625  
Db 563 QAPLNLNNTNNGTLLPSDPIGGKSNMFMNTLNTYGPLTALNNTAPVFPNGQIMDKELDT 622  
Qy 626 D--GHFHPSPLMGGFGLK-NPPQILIKNTPPVPANPFAEFS--TKFASFITQYSTGQSV 681  
Db 623 DLKPRLH---VTAPFVCNNPFGQLFVKIAP---NLTDENADSPOQPRIITYSNFWMKG 676  
Qy 682 EIEWELOKENSKRWNPEVOYTSNYAKSANVDFTVDNGLYTEPRPIG 728  
Db 677 TLFTAKRSSNMWNP IQOHTT-----TAENIGNYI-PTNIG 712

RESULT 7  
VCPVLP  
coat protein VP1 - feline panleukopenia virus  
N;Contains: coat protein VP2  
C;Species: feline panleukopenia virus, FPLV  
C;Date: 30-Jun-1987 #sequence\_revision 30-Jun-1987 #text\_change 16-Jul-1999  
C;Accession: A03701  
R;Carlson, J.; Rushlow, K.; Maxwell, I.; Maxwell, F.; Winston, S.; Hahn, W.  
J. Virol. 55, 574-587, 1985  
A;Title: Cloning and sequence of DNA encoding structural proteins of the autonomous parv  
A;Reference number: A03697; MUID:85265017; PMID:2991581  
A;Accession: A03701  
A;Molecule type: DNA  
A;Residues: 1-727 <CAR>  
A;Cross-references: EMBL:M10824; NID:g333474; PIDN:AAA47161.1; PID:g333476  
C;Genetics: 11/1  
C;Superfamily: parvovirus coat protein  
C;Keywords: coat protein  
F;144-727/Product: coat protein VP2 #status predicted <VP2>

Query Match 11.1%; Score 444.5; DB 1; Length 727;  
Best Local Similarity 23.7%; Pred. No. 3.7e-21;  
Matches 187; Conservative 117; Mismatches 295; Indels 189; Gaps 35;

Qy 43 RGLVLPGYKYLGPENGLDKGEPVNAADAALHDKAYDQOLKAGDNPYLRYNHADAEPQE 102  
Db 10 RGLVPPGYKYLGPENSLDGEPTNPSDAAKEHDEAYAAVYLRSGKNPYLYFSPADQRFID 69  
Qy 103 RLQEDTSFGNLRGAVFOAKKRVLEPLGLVEBGAKTAPGKRPEVQSPQEPDSSSGIGKT 162  
Db 70 QTKDATDWGKIGHYFFRAKALAPVLTDPDHPSTSRPTK-PTKRSKPPHIFINLAK- 127  
Qy 163 GOQPAKKRLNFGQTGDSSEVP---DPQPLGEPRA--TPAAVGPPTMASGGGAPMADNNE 216  
Db 128 -----KKKAGAGQVKRDNOAPMSDGAVQPDGGQPAVRNERATGSGNGSGGG--GG 176  
Qy 217 GADGVNAGNWH-----CDSTWLGDVRVITSTRTWALPTYNNHLYKOI-----SS 262  
Db 177 GSGGVGISTGTENNQTEFKFLENGVW--EITANSSRLVHLNMPSESENYKRVVNNMDKTA 234  
Qy 263 ASTGASNDNHYFGYSTPWGYFDENRFHCHFSPRDMQRLINNMWGFPRKRLNFKLENIOVK 322  
Db 235 VKGNMALDDTHVQIVTPWLSVDANAMGWENPGDWQLIVNTMSELHLVSEQEIFNVVLK 294  
Qy 323 EVT---TNDGVTIANNLSTVQVFSDEYQLPYVLGSAHQGL-----PPFPA----- 368

Db 295 TVSESATQPTKVYNNDLTASLWALDSNNTMFTPAAMRSETLGFYPMKPTIPTWRY 354
 Qy 369 ---DVEMIPQYGYLLTNGSQAVGRS-----SFYCLE-YFPSQMLRTGNNF-TFS 413
 Db 355 PQWDRTLIPSH-----TGTSGTPTNIYHGTDPDDVQFYTIENSVPVHLRTGDEFATGT 408
 Qy 414 YTFEEVPEHSSYAHQSQSLDRLMN--PLIDQYLYLNRTQNGSGSAQNKDLLFSRGSFAGM 471
 Db 409 FFFDCKP--CRLTHTWQTNRALGLRP-----FLNSLPQSEGATNFGDIGVQQDKRRGV 459
 Qy 472 S-----VQPKN-WLPGPCYRQQRVSK-----TKTDNNSNFTWT 504
 Db 460 TOMGNTDITTEATIMRPAEVGYSAFYSSFEASTQGPFKIPIAAGRGGAQTDENQAA--D 516
 Qy 505 GASKY--NLNGRESIINPGT---AMASHKDEDEKFFPMGVMIFGKESAGASNTALD- 556
 Db 517 GDPRYAFGRQHGQKTTTGTETPERFTYIAHQDT-----GRYPAGDWIQNINF 563
 Qy 557 NVMITDEEIKATNPVATERFGTVAVNFQSSSTDPAATGDVHANGALPGMWQDRDVTYLOG 616
 Db 564 NLPVTNDNVLLPTDPIG---GKTGINY--TNIENYGPLTALNNVP-----PVYPNG 610
 Qy 617 PIWAKIPHTDGHFHPSPLMGFGLK-----NPPQILIKNTVPAN--PPA 660
 Db 611 QIWDKEFDT-----LKPRLHVNAFVCCQNNCPGQLFVKVAPNLITNEYDPDA 657
 Qy 661 EFSATKFAFITQYSTGVSVLEIEMELQKENSKRWNPEVOYTSNYAKSANVDFTVDNNGL 720
 Db 658 SANMSR---IVTYSDFMWKGLVFKAKLRASHTWNPICQMSIN-----VDNQFN 703
 Qy 721 YTEPRPIG 728
 Db 704 YL-PNNIG 710

RESULT 8

VCPVME

coat protein VP1 - mink enteritis virus (strain Abashiri)
 N:Contains: coat protein VP2
 C:Species: mink enteritis virus, MEV
 C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 12-Apr-1996
 C:Accession: B38350
 R:Kariatsumari, T.; Horiuchi, M.; Hama, E.; Yaguchi, K.; Ishiguro, N.; Goto, H.; Shinag
 J. Gen. Virol. 72, 867-875, 1991
 A:Title: Construction and nucleotide sequence analysis of an infectious DNA clone of the
 A:Reference number: A38350; MUID:91202123; PMID:2016597
 A:Accession: B38350
 A:Molecule type: DNA
 A:Residues: 1-722 <KAR>
 A:Cross-references: GB:D00765
 C:Superfamily: parvovirus coat protein
 C:Keywords: coat protein
 F,139-722/Product: coat protein VP2 #status predicted <VP2>

Query Match 11.0%; Score 439.5; DB 1; Length 722;
 Best Local Similarity 23.6%; Pred. No. 7.7e-21;
 Matches 186; Conservative 117; Mismatches 295; Indels 189; Gaps 35;

Qy 44 GLVLPGYKYLGPFGNGLDKGEPVNAADAALHDKAYDQQLKAGDNPYLRYNHADAEPQER 103
 Db 6 GLVPPGYKYLGPFGNSLDQGEPTNPDAAKEHDEAYAYLRSGKNPYLYFSPADQRFIDQ 65
 Qy 104 LOEDTSPFGNLRGAVFQAKKRVLEPLGLVEGAKTAPGKKRPVEQSPQEPDSSSGIGKTG 163
 Db 66 TKDATDWGKIGHYFFRAKKAIAFVLTDTPDNPTSRPTK-PTKRSKPPPHIFINLAK-- 122
 Qy 164 QQPAKRLNFGQTGDSSEVP---DPQUGEPRA--TPAAVPTTMASSGGGAPMADNNEG 217
 Db 123 ---KKKAGAGQVCRDNLAPMSDGAVPDGGQPAVRNERATGSGNGSGGG---GCG 172
 Qy 218 ADGVGNASGNW-----CDSTWLGDRVITSTRTWALPTYNNHLYKQI-----SSA 263
 Db 173 SGGVGISTGTENNQTEFKPLENGWV--EITANSSRLVHLNMPESENYKRVVNNMDKTAV 230

Qy 264 STGASNDNHVFGYSTPWGYFDENRFHCHFSPRDWQRLINNNWGFPRKRLNFKLENIOYE 323
 Db 231 KGNMALDDTHVQIVTPMSLYDANAMGWENFGDWQLIVNTMSELHLVSFEQEIFNVILKT 290
 Qy 324 VT---TNDGVTTIANNLTSIVQVFSDBEYQLPYVLGSAHQCL-----PPFA----- 368
 Db 291 VSESATQPTKVYNNDLTASLWALDSNNTMFTPAAMRSETLGFYPMKPTIPTWRY 350
 Qy 369 --DVEMIPQYGYLLTNGSQAVGRS-----SFYCLE-YFPSQMLRTGNNF-TFSY 414
 Db 351 QWDRTLIPSH-----TGTSGTPTNIYHGTDPDDVQFYTIENSVPVHLRTGDEFATGT 404
 Qy 415 TFEVPEHSSYAHQSQSLDRLMN--PLIDQYLYLNRTQNGSGSAQNKDLLFSRGSFAGMS 472
 Db 405 FFFDCKP--CRLTHTWQTNRALGLRP-----FLNSLPQSEGATNFGDIGVQQDKRRGV 455
 Qy 473 -----VQPKN-WLPGPCYRQQRVSK-----TKTDNNSNFTWTG 505
 Db 456 QMGNTDITTEATIMRPAEVGYSAFYSSFEASTQGPFKIPIAAGRGGAQTDENQAA---DG 512
 Qy 506 ASKY--NLNGRESIINPGT---AMASHKDEDEKFFPMGVMIFGKESAGASNTALD-N 557
 Db 513 DPRYAFGRQHGQKTTTGTETPERFTYIAHQDT-----GRYPAGDWIQNINF 559
 Qy 558 VMITDEEIKATNPVATERFGTVAVNFQSSSTDPAATGDVHANGALPGMWQDRDVTYLOG 617
 Db 560 LPVTNDNVLLPTDPIG---GKTGINY--TNIENYGPLTALNNVP-----PVYPNGQ 606
 Qy 618 IWAKIPHTDGHFHPSPLMGFGLK-----NPPQILIKNTVPAN--PPA 661
 Db 607 IWDKEFDT-----LKPRLHVNAFVCCQNNCPGQLFVKVAPNLITNEYDPDAS 653
 Qy 662 FSATKFAFITQYSTGVSVLEIEMELQKENSKRWNPEVOYTSNYAKSANVDFTVDNNGLY 721
 Db 654 ANMSR---IVTYSDFMWKGLVFKAKLRASHTWNPICQMSIN-----VDNQFN 699
 Qy 722 TEPRPIG 728
 Db 700 L-PNNIG 705

RESULT 9

VCPVFP

coat protein VP1 - feline panleukopenia virus (strain 193)
 N:Contains: coat protein VP2
 C:Species: feline panleukopenia virus, FPLV
 C:Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 12-Apr-1996
 C:Accession: B36608
 R:Martyn, J.C.; Davidson, B.E.; Studdert, M.J.
 J. Gen. Virol. 71, 2747-2753, 1990
 A:Title: Nucleotide sequence of feline panleukopenia virus: comparison with canine parvo
 A:Reference number: A36608; MUID:91073139; PMID:2174965
 A:Accession: B36608
 A:Molecule type: DNA
 A:Residues: 1-727 <MAR>
 A:Cross-references: GB:X55115
 C:Superfamily: parvovirus coat protein
 C:Keywords: coat protein
 F,144-727/Product: coat protein VP2 #status predicted <VP2>

Query Match 10.8%; Score 431.5; DB 1; Length 727;
 Best Local Similarity 23.5%; Pred. No. 2.6e-20;
 Matches 185; Conservative 117; Mismatches 297; Indels 189; Gaps 35;

Qy 43 RGLVLPGYKYLGPFGNGLDKGEPVNAADAALHDKAYDQQLKAGDNPYLRYNHADAEPQER 102
 Db 10 RGLVPPGYKYLGPFGNSLDQGEPTNPDAAKEHDEAYAYLRSGKNPYLYFSPADQRFID 69
 Qy 103 RLOEDTSPFGNLRGAVFQAKKRVLEPLGLVEGAKTAPGKKRPVEQSPQEPDSSSGIGKT 162
 Db 70 QTKDAKDWGKIGHYFFRAKKAIAFVLTDTPDHPSTSRPTK-PTKRSKPPPHIFINLAK- 127

QY	163	GOQPAKRLNFGQTDSESV	----	DPQPLGEPPA--	TPAAV3PTT	MASGGGAP	ADNNE	216
		:						
Db	128	----	KKKAGAGQV	KRDNLA	PMSDGA	VQPDGG	PAVRNERAT	SGSGGGG
		:						
QY	217	GADGVNASGNWH	-----	CDSTWLG	RVI	TTSTR	TWALPT	YNNHLYKQI
		:						
Db	177	GSQGVG	ISTGTENNQTE	EKFL	ENGWV--	EI	TRANSRLV	HLNMPES
		:						
QY	263	ASTGASNDNH	YFGYST	PMGY	FDNRF	CHFS	PRDQRL	INNNGFR
		:						
Db	235	VKGNMAL	DDIHVQI	VT	PSL	VDANAM	GWFWEN	PBGWQI
		:						
QY	323	EVT	----	TNDG	VTTIAN	LSTVQV	FS	DEYQLPY
		:						
Db	295	TVSE	SATQPP	TKVYNN	DLTAS	LMALDS	NN	TMPFTPA
		:						
QY	369	--	DVFMIPQY	GLTL	NNGSQ	AVGRS	-----	SFYCLE
		:						
Db	355	PQWDR	TLPSH	----	TGTS	GPTPNV	HGTB	DDVQFY
		:						
QY	414	YTFEEV	PEHSSYA	HSQSL	DL	LMN--	PLIDQ	YLYLN
		:						
Db	409	FFFDCKP	--	CR	LHTWQ	TNRAL	GLRP	-----
		:						
QY	472	S	-----	VQPKN	WL	PGPCY	RQQRVSK	-----
		:						
Db	460	TQMG	NTDYIT	EATIM	RPAE	VGYSAP	YSSF	EASTQGP
		:						
QY	505	GASKY	----	NLNG	RESI	IN	PGT---	AMASHK
		:						
Db	517	GDP	RYAFGRQ	HQK	TTT	GET	PER	FTYIAHQDT
		:						
QY	557	NVMT	DEE	EIKAT	NP	VATER	FGT	VAVNFQSS
		:						
Db	564	NLP	VTND	NVLL	PTD	PIG---	GKT	GINY--
		:						
QY	617	PIMAKI	PH	TDGH	FH	PS	PLMG	GLK-----
		:						
Db	611	QI	WKE	EDTD	-----	LK	PR	LHVNA
		:						
QY	661	EFS	ATK	FAS	FITQ	YSTG	QVSEI	EWELO
		:						
Db	658	SA	NMSR	----	IV	TS	D	FWWKGL
		:						
QY	721	Y	TE	PR	P	I	G	728
		:						
Db	704	YV	-	P	N	N	I	G 710
		:						

```

RESULT 10
VCPVPP
coat protein VP1 - porcine parvovirus
N;Contains: coat protein VP2
C;Species: porcine parvovirus
C;Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 12-Apr-1996
C;Accession: B33302
R;Ranz, A.I.; Mancius, J.J.; Diaz-Aroca, E.; Casal, J.I.
J. Gen. Virol. 70, 2541-2553, 1989
A;Title: Porcine parvovirus: DNA sequence and genome organization.
A;Reference number: A33302; MUID:90010964; PMID:2794971
A;Accession: B33302
A;Molecule type: DNA
A;Residues: 1-723 <RAN>
A;Cross-references: EMBL:D00623
C;Genetics:
A;Introns: 10/1
C;Superfamily: parvovirus coat protein
C;Keywords: coat protein
F;145-723/Product: coat protein VP2 #status predicted <VP2>

```

Query Match	10.7%;	Score 426;	DB 1;	Length 723;
Best Local Similarity	24.9%;	Pred. No. 5.9e-20;		
Matches 190;	Conservative 109;	Mismatches 308;	Indels 156;	Gaps 36;

[illegible]

```

RESULT 11
VCPVIM
coat|protein VP1 - minute virus of mice (strain MVMi)
C;Species: minute virus of mice, murine parvovirus
C;Date: 28-Dec-1987 #sequence_revision 28-Dec-1987 #text_change 28-Jul-2000
C;Accession: B23008; B29510
R;Sahl, R.; McMaster, G.K.; Hirt, B.
Nucleic Acids Res. 13, 3617-3633, 1985
A;Title: DNA sequence comparison between two tissue-specific variants of the autonomous I
A;Reference number: A23008; MUID:85242059; PMID:3855242
A;Accession: B23008
A;Molecule type: DNA
A;Residues: 1-718 <SAH>
A;Cross-references: EMBL:X02481; NID:G60918; PIDN:CAB46507.1; PID:G5419928
R;Astell, C.R.; Gardiner, E.M.; Tattersall, P.
J. Virol. 57, 656-669, 1986
A;Title: DNA sequence of the lymphotropic variant of minute virus of mice, MVM(1), and c
A;Reference number: A29510; MUID:86115415; PMID:3502703
A;Accession: B29510

```





VCPVCD  
 coat protein VP1 - canine parvovirus (strain CPV-d)  
 N;Contains: coat protein VP2  
 C;Species: canine parvovirus, CPV  
 C;Date: 31-Mar-1990 #sequence\_revision 31-Mar-1990 #text\_change 16-Jul-1999  
 C;Accession: A31163  
 R;Parish, C.R.; Aquadro, C.F.; Carmichael, L.E.  
 Virolgy 166, 293-307, 1988  
 A;Title: Canine host range and a specific epitope map along with variant sequences in the  
 A;Reference number: A31163, MUID:89020796, PMID:3176341  
 A;Accession: A31163  
 A;Molecule type: DNA  
 A;Residues: 1-737 <PAR>  
 A;Cross-references: EMBL:M23255, NID:g333467, PIDN:AAA47158.1, PID:g333468  
 C;Genetics:  
 A;Introns: 26/3  
 C;Superfamily: parvovirus coat protein  
 C;Keywords: coat protein  
 F;584-737/Product: coat protein VP2 #status predicted <VP2>

Query Match	10.2%;	Score 407.5;	DB 1;	Length 737;
Best Local Similarity	22.8%;	Pred. No. 9.8e-19;		
Matches 181;	Conservative 112;	Mismatches 301;	Indels 201;	Gaps 32;

[illegible]

Db	661	NEYDEPDASANMSR---	IVTYSDFEWKKGKLVFAKLRASH	TWNP	IQOM	SIN-----	707
QY	714	TVDDNGLYTEPRPIG	728				
Db	708	-VDNQFNIV-PSNIG	720				

RESULT 14  
VCPVW2

coat protein VP1 - parvovirus H1

C;Species: parvovirus H1

A;Note: host Homo sapiens (man)

C;Date: 14-Nov-1983 #sequence revision 14-Nov-1983 #text\_change 08-Apr-1994

C;Accession: A03699

R;Rhode III, S.L.; Paradiso, P.R.

J. VIROL. 45, 173-184, 1983

A:Title: Parvovirus genome: nucleotide sequence of H-1

A:Reference number:

A: Accession: A03699  
A: Reference number: AV3000, HOLD: 00112100, 2

```
A:MOLECULE type: DN
A:ACCESSION: A03633
```

A:Residues: 1-722 <RHO>  
A:Molecule type: DNA

A;Cros8-references: EMBL:X01457: EMBL:J02198  
A;residues: 1-722 <KHO>

C: Superfamily: Daryovirine coat protein

C:superfamily: parvoviru  
C:keywords: coat protein

C;Keywords: coat protein

Query Match	10.1%;	Score 404;	DB 1;	length 722;
Best Local Similarity	23.3%;	Pred. No. 1.6e-18;		
Matches 175;	Conservative 108;	Mismatches 290;	Indels 178;	Gaps 30;

**DQ** 57 NGLDKGEIPVNAADAAL EHKDAYDQULKAGDNPLYLRYNHADAEFQERLOEDTSPFGNLGR 116  
| | : | | | : | | | | : | | | : | | : | | : | | : | | : | | : | | : | | :  
**Dd** 12 NSLDGEEPTNPSDAAKEKHDEAYDYIKSGKNPYLYFSPADORFI DQTCKADKWGGKVGH 71

117 AVFOAKRRVLEPLGLVEEGAKTAPGKKRPVEQSPOEPDSSSGIGKTGQQ---PA----- 167

```

      : : | | : : : : : :
72 YFRTKRAFAPKLTDS-----PG-----TSGVSRPGKRTKPPAHIFVNO 112
Db

```

168 ----KKRLNFGGTGDSSEVPDPPLTGEPATPAVGPPTMASGGGAPMADNNEGADGVGN 223

Db 113 ARAKKRASLAQORTLTMSDGETINQPTGTIANARVERSADGGGS -SGGGGSGGGGIGV 171

```

QY      224 ASGNHCDSTW--LGDRVITSTRVAL-----PTYNNHLYKOISSASTGA 267
      ::||: :||| : : : :||| :

```

Db 172 STGTYYNQTTKKFLGDGWVEITAAASRLLIHGMPPSENYCRVTYHNNQTTGHTKVKGNM 231

[illegible]

DB	232	AYDTHQIWM-IPMSLVDANAMGVMQPSDMMQFIQNSMESLNLJLSLSQLEPNVVAIVIEQ	290
DB	327	NDCVETIANNITESTIVFSDSEYOLPVYLSAHOGL-----BPBPAD---YEMT	373

291	CGAGGDAIKYNNNDITACMMVAALDSINNILPYTPAAOTSETTIGFYPWKPAPABARYFEEM	350
dy	NDVSHHFNANLISIVVIVDSSELYFTVDSNNGSC	410
dy	NDVSHHFNANLISIVVIVDSSELYFTVDSNNGSC	410

374 PORGYLINNGSQ-----AVG-----RSSFYCLE-YFPSOMLRTGNFTF-SYTFEEVP 420

Db 351 PROI:SVTSSNSAEGTQITDTIGEPQALNSQFTIENTLPTLLRLTGDEFTTGYIENIDP 410

421 FHSSYA-----HSQSLDRLMN-PLIDQYLYYLNRTQNGSGAQNKKLLF----- 463 QY

411 LKTHWQNRHLACLOGITDLPISITATASLTANGDRFGSTQTONVNYVEALRTRPAQ 470

464 -----SRGSPAGMSVQPKNWLPGPCYRQORVSKTKTDNNNSNFTWGASKYNL 511  
 .||| . :|||

Db 471 IGENPHDNFEANRGPFKVPVP-----LDITAGEDHDAN-----GAIRFNY 513

QY 512 NGR--ESIINPGTA-----MASHKODEDEKFFPMGVMIFCKESAGASNTALDNVM 559

Db 514 GKQHGEDWAKOGAPERYTWDALDSAGRDTRCFV-----QSAPISIPNONQI 563

```

QY      560 ITDDEEIK- TNPVATERFGTVAANFQSSSTDPA TGDVHAMGALPGMWQZBRDYLQGP I 618
      : - : | | | | : | | |
Db      564 LQREDAIAGRTNMHYTNVFNYSGPLSAFPHDP-----IYPNGQI 603

```

Qy 619 WAK--IPHTDGHFHPSPLMGFGGLKNPPQILLIKNTVPANPPAEF--SATKFASFITQ 673  
Db 604 WDKELDLDEHKERLHVTAPEFV---CKNNPQGLFVHLGP---NLTDQFDPNSTTVSRIVT- 656  
Qy 674 YSTGQVSVEIEWELQKENSKRWNPEVOYTSN 704  
Db 657 YSTFYWKGLKFKAKLRPNLTWNPFVQATTD 687

RESULT 15  
VCPVCN

coat protein VP1 - canine parvovirus (strain 780929)  
N:Contains: coat protein VP2  
C:Species: canine parvovirus, CPV  
C:Date: 30-Jun-1987 #sequence\_revision 30-Jun-1987 #text\_change 12-Apr-1996  
C:Accession: A03702  
R:Rhode III, S.L.  
J. Virol. 54, 630-633, 1985  
A:Title: Nucleotide sequence of the coat protein gene of canine parvovirus.  
A:Reference number: A03702; MUID:85185696; PMID:3989914  
A:Accession: A03702  
A:Molecule type: DNA  
A:Residues: 1-722 <RHO>  
A:Cross-references: EMBL:M10989  
C:Genetics:  
A:Introns: 11/1  
C:Superfamily: parvovirus coat protein  
C:Keywords: coat protein  
F:139-722/Product: coat protein VP2 #status predicted <VP2>

Query Match 9.7%; Score 385.5; DB 1; Length 722;  
Best Local Similarity 22.8%; Pred. No. 2.6e-17;  
Matches 178; Conservative 118; Mismatches 297; Indels 189; Gaps 35;

Qy 49 GYKYLGFENGLDKGEFVNADAALAEHDKAYDQQLKAGDNPLYRYNHADAEFQERLQEDT 108  
Db 11 GYKYLGPCKSLDQGEPTNPSDAAKEHDEAYAYLRSGKNPLYFSPADQRFIDQTDKAK 70  
Qy 109 SFGNLRGAVFOAKKRVLEPLGLVEEAKTAPGKKRPVEQSPQEPDSSSGIGTGQOPAK 168  
Db 71 DWGKIGHYFFRAKKAIAIPVLTDPDHPSTSRPTK-PTKRSKPPPHIFINLAK-----K 123  
Qy 169 KRLNFGQTGDSSEVP---DPQPLGEPRA--TPAVGPTTMASGGGAPMADNNEGADVG 222  
Db 124 KKAGAGQYKRDNLAPMSDGAQPDGQDAVRNERATSGSGGGG-----GGSGGCVG 177  
Qy 223 NASGNW-----CDSTWLGDRVITSTRTWALPTYNNHLYKQI-----SSASTGAS 268  
Db 178 ISTGFNNQTEFKLENGWV--EITANSSRLVHLNMPSESEKDRRYVNNMDKTAVNGNMA 235  
Qy 269 NDNHYFGYSTPWGYFDNRFHCHFSPPDWQRLINNNGFRRPKRLNFKLENIOVKEVT--- 325  
Db 236 LDDIHAQIVTPWSLVDANAWDWENFGDWQLIVNTMSEHLVSEFQELFNVLKTVSESA 295  
Qy 326 TNDGVTTIANNLTSTVQVFSDESEYQLPYVLGSAHQGL-----PPFPA-----DVF 371  
Db 296 TQPTKYVNNDLTASLWALDSNNMTFTPAARSETLGFYPMKPTIIPWRYYPQWDRT 355  
Qy 372 MIPQYGLTLNNGSAVGRS-----SFYCLE-YFPQMLRTGNNE-TFSYTFEEV 419  
Db 356 LIPSH-----TGTSGLPTNIYHGTDDVDVQFYTIENSVPHLRTGDEFATGTFEFDCK 409  
Qy 420 PFHSSYAHQSQSLDRLMNPIDQYLYINRTQNS-----GSAQNK-----D 460  
Db 410 P--CRLTHWTQTNRALG--LPPLNSLPSQSEGATNFGDIGVPDCKKRGVTOGNTNYITE 465  
Qy 461 LIFSRSPPAGMSVQ-----PKMWLPGPCYRQQRVSKTKTDNNNSNFTWTGASKY- 509  
Db 466 ATIMRPAEYGYSAFYVSFEASTQGFKTLPIAAGR-----GGAQTDENQA-----ADGNRYA 517  
Qy 510 --NLNGRESLINPCT---AMASHKDEDEKFFPMGVMIFGKESAGASNTALD-NVMITD 562

Db 518 FGRQHGKTTTGTGETPERFTYIAHQDT-----GRYPEGDWIONINENLBYTN 564  
Qy 563 EBEIKATNPVATERFGTAVAVNFQSSSTDPAATGVHAMGALPGMWODRDVYLOGPIWAKI 622  
Db 565 DNVLPIPIDIG---GKTGINV--TNIFNTYGPPLTALNNVP-----PVYPNGQIWDKE 611  
Qy 623 PHTDGHFHPSPLMGFGGLK-----NPPQILLIKNTVPAN---PPAEFSATK 666  
Db 612 FDTD-----LKPRLVNAPFVCCNNCPGQLFVKLAPNLTNEYDPDASANMSR 658  
Qy 667 FASFTQYSTGQVSVEIEWELQKENSKRWNPEVOYTSNYAKSANVDFTVDNNGLYTEBRP 726  
Db 659 ----IVTYSDFWWMKGKLVFKAKLRASHWTWNPICQMSIN-----VDNQFNIV-PSN 703  
Qy 727 IG 728  
Db 704 IG 705

Search completed: July 17, 2003, 18:35:44  
Job time : 30.3842 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 2003, 18:20:49 ; Search time 14.5704 Seconds  
(without alignments)  
2095.115 Million cell updates/sec

Title: US-09-807-802A-13  
Perfect score: 3989  
Sequence: 1 MAADGYLPDWLEMDNLSEGIR.....NNGLYTEPRPIGTRYLLTRPL 736

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2177.5	54.6	504	1 COA3_AAV2	P03135 adeno-ascoc
2	798.5	20.0	673	1 COAT_PAVBO	P07297 bovine parv
3	600.5	15.1	781	1 COAT_PAVHB	P07299 human parvo
4	482	12.1	729	1 COA1_PAVP9	P33484 porcine par
5	482	12.1	729	1 COA1_PAVP9	P18546 porcine par
6	480	12.0	729	1 COA1_PAVPK	P52501 porcine par
7	444.5	11.1	727	1 COAT_FPV	P04864 feline panl
8	441.5	11.1	749	1 COA2_PAVPN	P22964 porcine par
9	437.5	11.0	722	1 COAT_MEVA	P27437 mink enteri
10	431.5	10.8	727	1 COAT_FPV19	P24840 feline panl
11	430.5	10.6	718	1 COAT_PAVCB	P11213 canine parv
12	424.5	10.6	718	1 COAT_MDMIM	P07302 murine minu
13	421.5	10.6	748	1 COAT_PAVCN	P12930 canine parv
14	407.5	10.2	737	1 COAT_PAVCD	P17455 canine parv
15	404	10.1	722	1 COAT_PAVHH	P03136 hamster par
16	386.5	9.7	722	1 COAT_PAVC7	P04863 canine parv
17	377	9.5	716	1 COAT_MDMIV	P03137 murine minu
18	234	5.9	587	1 COAT_PAVL3	P36310 parvovirus
19	207.5	5.2	809	1 COAT_DSDDV	O71155 diatraea sa
20	201	5.0	811	1 COAT_GMDNV	O90125 gallieria me
21	199.5	5.0	810	1 COAT_JCDNV	O90053 junonia coe
22	195	4.9	584	1 COAT_PAVC2	P30129 canine parv
23	186.5	4.7	647	1 COAT_PAVG	P24029 aleutian mi
24	143.5	3.6	648	1 WHN_MOUSE	O61575 mus musculu
25	131.5	3.3	1446	1 IE18_PRYKA	P33479 pseudorabie
26	126.5	3.2	1461	1 IE18_PRYTF	P1675 pseudorabie
27	124.5	3.1	2493	1 CYAA_USUMA	P49606 ustilago ma
28	123	3.1	648	1 WHN_HUMAN	O15353 homo sapien
29	123	3.1	1142	1 ENAM_PIG	O97939 sus scrofa
30	121	3.0	880	1 SYV_BACST	P11931 bacillus st
31	121	3.0	1113	1 YFF9_YEAST	O02630 saccharomyc
32	120	3.0	1379	1 YFF9_SCHPO	O14066 schizosacch
33	119.5	3.0	1394	1 HAP_HAEIN	P45387 haemophilus

34	119.5	3.0	2004	1 YP73_YERPE	O8zdj2 yersinia pe
35	119	3.0	667	1 ITR2_HUMAN	P15884 homo sapien
36	117.5	2.9	1076	1 NUP1_YEAST	P20676 saccharomyc
37	116.5	2.9	642	1 ITR2_CANFA	P15881 canis famill
38	114.5	2.9	802	1 PAC_BACME	O60136 bacillus me
39	113.5	2.8	1324	1 IRS2_HUMAN	O9y4h2 homo sapien
40	113	2.8	837	1 HFC2_HAEIN	P45997 haemophilus
41	113	2.8	1484	1 CES2_HUMAN	O9bxf3 homo sapien
42	112	2.8	774	1 FECA_ECOLI	P13036 escherichia
43	111.5	2.8	2944	1 CAL7_HUMAN	O02388 homo sapien
44	111	2.8	454	1 NOE2_HUMAN	O95897 homo sapien
45	111	2.8	635	1 HMLA_DROME	P10105 drosophila

ALIGNMENTS

RESULT 1	ID	COA3_AAV2	STANDARD;	PRT;	504 AA.
AC	P03135;				
DT	21-JUL-1986 (Rel. 01, Created)				
DT	21-JUL-1986 (Rel. 01, Last sequence update)				
DT	01-NOV-1991 (Rel. 20, Last annotation update)				
DE	Probable coat protein 3.				
OS	Adeno-associated virus 2 (AAV2).				
OC	Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.				
OX	NCBI_TaxID=10804;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=83164299; Pubmed=6300419;				
RA	Srivastava A., Iusby E.W., Berns K.L.;				
RT	"Nucleotide sequence and organization of the adeno-associated virus 2 genome.";				
RL	J. Virol. 45:555-564(1983).				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL; J01901; AAA42376.1; --				
DR	PIR; A03698; VCPV3A.				
DR	InterPro; IPR001403; Parvo_coat.				
DR	Pfam; PF00740; Parvo_coat; 1.				
KW	Coat protein.				
SQ	SEQUENCE 504 AA; 56366 MW; 758999B017052B6B2 CRC64;				
QY	Query Match	54.6%;	Score 2177.5;	DB 1;	Length 504;
QY	Best Local Similarity	82.5%;	Pred. No. 2.5e-128;		
QY	Matches 392;	Conservative 33;	Mismatches 47;	Indels 3;	Gaps 2;
QY	203 MASGGAPMADNNEGADGVGNASGNWCHDSTWLGDRVITTTSTRTWALPTYNHLYKQISS				262
Db	1 MATSGAPMADNNEGADGVGNSSGNWCHDSTWMDRVITTTSTRTWALPTYNHLYKQISS				60
QY	263 ASTGASNDNHFGYSTPWGYFDENRFCHFSPRDWQRLINNNGWGRPKRLNFKLFNIQVK				322
Db	61 QS-GASNDNHFGYSTPWGYFDENRFCHFSPRDWQRLINNNGWGRPKRLNFKLFNIQVK				119
QY	323 EYTTNDGVTTIANNLTSTVOFSSDSEYQLPYVLGSAHOGCLPPFPADVFMIPQYGYLTN				382
Db	120 EYTONDGTITIANNLSTFVQYFTDSEYQLPYVLGSAHOGCLPPFPADVFMIPQYGYLTN				179
QY	383 NGSQAVGRSSFYCLEYFPQMLRTGNNFTFSYTFEEVPEFSSYAHQSQSLDRLMPLIDQY				442
Db	180 NGSQAVGRSSFYCLEYFPQMLRTGNNFTFSYTFEDVPEFSSYAHQSQSLDRLMPLIDQY				239
QY	443 LYTLNRTONQSGAQNKDLFSRGSPPAGMSVQPKNMLPQPCYRQQRVSKTKTDNNNSNFT				502

Db 240 LYYLSRTNTPSGTTTQSRLQFSQAGASDIRDSRNWLPGPCYRQQRVSKTSADNNNSEYS 299  
QY 503 WTGASKYNLNGRESIINPGTAMASHKODEDEKFFPMGVMIFGKESAGASNTALDNVMTD 562  
Db 300 WTGATKYHLNGRDLVNP--AMASHKODEEKFFPQSGVLIFFKQSGSEKTNVIEKMITD 357  
QY 563 EEEIKATNPVATERGTVAVNFQSSSTDPAAGVHAMGALPGMWQDRDLYLQGPIMAKI 622  
Db 358 EEEIGTNPVATEGYGSVSTNLQGRNRQAATADVNTQGVLPGMWQDRDLYLQGPIMAKI 417  
QY 623 PHTDGHFHPSPLMGFGFLKNPPQILIKNTPVANPPAEFSATKFASTTQYSTG 677  
Db 418 PHTDGHFHPSPLMGFGFLKNPPQILIKNTPVANPPSTTFSAKFASTTQYSTG 472

## RESULT 2

COAT\_PAVBO STANDARD; PRT; 673 AA.  
ID COAT\_PAVBO STANDARD; PRT; 673 AA.  
AC P07297; Q84374;  
DT 01-APR-1988 (Rel. 07, Created)  
DT 01-APR-1988 (Rel. 07, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Coat protein VP2 [Contains: Coat protein VP3].  
OS Bovine parvovirus (BPV).  
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.  
OX NCBI\_TaxID=10784;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=87061184; PubMed=3783814;  
RA Chen K.C., Shull B.C., Moses E.A., Lederman M., Stout E.R.,  
RA Bates R.C.;  
RT "Complete nucleotide sequence and genome organization of bovine  
parvovirus.";  
RL J. Virol. 60:1085-1097(1986).  
CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M14363; AAB59848.1; -;  
DR EMBL; M14363; AAB59849.1; -;  
DR PIR; A26104; VCPVB5.  
DR HSSP; P30129; 4DPV.  
DR InterPro; IPR001403; Parvo\_coat.  
DR Pfam; PF00740; Parvo\_coat; 1.  
KW Coat protein; Glycoprotein.  
FT CHAIN 1 673 COAT PROTEIN VP2.  
FT CHAIN 138 673 COAT PROTEIN VP3.  
FT CARBOHYD 343 343 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 344 344 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 446 446 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 639 639 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT DOMAIN 163 187 GLY-RICH.  
SQ SEQUENCE 673 AA; 75103 MW; 5F244642B2214831 CRC64;

Query Match 20.0%; Score 798.5; DB 1; Length 673;  
Best Local Similarity 28.6%; Pred. No. 2.1e-42;  
Matches 214; Conservative 105; Mismatches 275; Indels 155; Gaps 21;

QY 30 PKPKANQKQDDGRGLVLPYKYLGPFNGLDKGEVNAADAALAEHDKAYDQOLKAGDNP 89  
Db 3 PTNKANSKK-----GLTLPGYNYLGPNSLFGAPVNKADAAARKHDFGYSDDLKEGKNP 57  
QY 90 YLRYNHADAEPQERLQEDTSFGGNLGRAVPQAKRVLLEPLGLVEBGAATAPGKKRPVEQS 149  
Db 58 YLYFNTHDQNLIDELKDDTSFGGKLARGVFQIKALAPAL-----PGTSKGGDRALKRK 111  
QY 150 PQEPDSSSGIGKTGQOPAKKRINFGQTGDSSEVPDPQLGEP--PATPAVGPPTMASGG 207

Db 112 LYFARSNKGAKKANREPAPSTSNQONMEVSNDI PNDEAGNPIELATRSVVGSGVGCGG 171  
QY 208 GAPMADNNEGADGVGNASGNWHDSTWLGDVITTSRTTALPTYNHLYKQISSASTGA 267  
Db 172 -----RGSGVGYSTGCGTGGTIFSENI VVTKTNTQFICDIKNGHLYKS-EVLNTGD 222  
QY 268 SNDNHFGYSTPMGYFDFNRFHCHSPRDWRLINNNMGFRPKLNFNIQKEVTN 327  
Db 223 TAHRQY-AITPWSYFNFNQYSSHSPNDWQHLVNDYERFRPKAMIVRVNLIQIKQIMTD 281  
QY 328 DGVTTIANN-LTSTQVFSDSXQLPVYLSAHQGLPPPADVEMIPQGYLTL----- 381  
Db 282 GAMGTAVNNDLTAGMHI FCDGHRYPVQHPWDDQCMPELNSIWELPQYAYIPAPISVV 341  
QY 382 -NNGSQAVGRS-----SPYCLEYFPQSQMLRTGNFTFSYTFEEVFPHSSYAHQSGLDRM 435  
Db 342 DNNTFTNYEEHLKGVPLYLENSDHEVL RNG-----RIV 376  
QY 436 NPLIDQLYLYLNTQN-----QSGSAQNKDLFSRGP-----AGMSVQPKNWLP 480  
Db 377 RIVQLRLMDRKQHHIQHASDVQSTGQKQKLLIQRTQPNKQRFQNALRTSNMS 436  
QY 481 GPCYRQQRVSKTKTDNNNSNFTWTGASKYNLNGRESIINPGTAMASHKODEKFFPMGCV 540  
Db 437 GP-----GIARGTHNATLQTSAGALVTMTNGAD---VSGV 470  
QY 541 M-----IFGKE-----SAGASNTALDNVMTDEEIKATNPVATE 575  
Db 471 RAVRVGYSTDPITYGGQQPESDILRLRYASAAEQQNPILEN-----AARH 516  
QY 576 RFGTVAVNFQSSSTDPAAGVHAMGALPGMWQDRDLYLQGPIMAKIPHTDGHFHPSPLM 635  
Db 517 TTTREARTKLITGSGADGDYKEWMLPNQWMSAPISRYNPDIWKVPRVRNKTLLDTQD 576  
QY 636 GGFGLKNPPQILIK--NTPVANPPAEFSATKFASTTQYSTGQVSEIEMELQKENSX 693  
Db 577 GSIPMSHPGTLIFIKLARIPVPGND-----SFLNIYVTGQVSCVWEVEKEKGTG 627  
QY 694 RMNPEVOYTSNYAKSAND-FTVDNNGLY 721  
Db 628 NWRPEYVHS--ATNMSVDATYTNAGVY 653

## RESULT 3

COAT\_PAVHB STANDARD; PRT; 781 AA.  
ID COAT\_PAVHB STANDARD; PRT; 781 AA.  
AC P07297;  
DT 01-APR-1988 (Rel. 07, Created)  
DT 01-APR-1988 (Rel. 07, Last sequence update)  
DT 01-APR-1990 (Rel. 14, Last annotation update)  
DE Probable coat protein VP1.  
OS Human parvovirus B19.  
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Erythrovirus.  
OX NCBI\_TaxID=10798;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Isolate AU;  
RX MEDLINE=86200451; PubMed=3701931;  
RA Shade R.O., Blundell M.C., Cotmore S.F., Tattersall P., Astell C.R.;  
RT "Nucleotide sequence and genome organization of human parvovirus B19  
isolated from the serum of a child during aplastic crisis.";  
RL J. Virol. 58:921-936(1986).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M13178; AAA66867.1; -;



DR PIR; A24299; VCPV19.  
DR InterPro; IPR001403; Parvo\_coat.  
DR Pfam; PF00740; Parvo\_coat; 1.  
KW Coat protein; Glycoprotein.  
FT CARBOHYD 46 46 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 184 184 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 220 220 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 293 293 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 781 AA; 86015 MW; 8C6254DBD0576B07 CRC64;

Query Match 15.1%; Score 600.5; DB 1; Length 781;  
Best Local Similarity 26.6%; Pred. No. 5.3e-30;  
Matches 182; Conservative 95; Mismatches 300; Indels 107; Gaps 20;

QY 47 LPGYKYLGPFGNGLDKGEPVNAADAALAEHDKAYDQOLKAGDNPYLRYNHADAEFOERLOE 106  
DB 125 LPTGNYVGPENLQAGPPQSAVDSARIDFRYSQIAKLGINPYTHWTVADEELLKNIKN 184  
QY 107 DTSEGNLGRAVFOAKKRVLEPLGLVEEAKTAPGKKRPVEQSPQEPDSSSGIGKTGOOP 166  
DB 185 ETGFQAOVVKDYF-----TLKGAAPVAHFQ-----GSLPEVP 217  
QY 167 AKKRLNFGOTGDSSEVPDPPLGEPBPATPAVGPPTMASGGAPMADNNEGADGVGNASG 226  
DB 218 AYNA-----SEKYPMSVTSVNSAEASTGA-----GGG-----GSNSVKS 250  
QY 227 NMHCSTWLGDRTVITSTRTWALPTYYNNHLKYQISSASTGASND-----NHYFGY 276  
DB 251 NMSEGATFSANSVCTCFSRQFLIPYDEHHYKVFSPAASSCHNASGKEAKVCTISPIMGY 310  
QY 277 STEPGYFDENRHFCHFSRPDMQRLINNNGFRPKRLNFKLFNIQVKEVT--TNDGVTTIA 334  
DB 311 STPMRYLDENALNLFSPLEFQHLIENYGSIAPDALTVTISIAVKDVTDKTGGGV-QVT 369  
QY 335 NNLTSTVQVFSDESEYQLPYVLGSAHQGLPFPADVFMIPOGYVLTINN-GSQAVG---- 389  
DB 370 DSTGRLCMLVDHEYKYPYVLGGQDTLAPELPIWVYFPPQYAVLTWGDVNTQGISGDSK 429  
QY 390 -----RSSFYCLEYFSPSOMLRTGNNFTSYTFEEVFPFHSSVYAHQSOLDRLNPLIDQVLY 444  
DB 430 KLAEESEAFYVLEHSSFQLLGTGTASMSYKFPVPPEENLEGGSGHFYEMYNPL---YGS 486  
QY 445 YLNRTQNGSGAQNKDLLFSRGSFAGMSVQPKNMLPGPCYRQQRVSKTKTDNNNSNFTWT 504  
DB 487 RLGVPTDLGDDPKFRSL-----THEHAIQPNFMFPLVNSVSTKEGDSNTGAGKALT 541  
QY 505 GASKYNLNGRESIINPG-TAMASHKDEDEKFFPMSCVMIFGKESAGASNTALDNV----- 558  
DB 542 GLSTGTSQNTIRISLRPGPVSOPIHHWDTKYVTGINAISHGQTTYG---NAEDKEYQGV 598  
QY 559 --MITDEEIKATNPVATERFGTAVAVNFQSSSTDPATGVDHAMGALPGMWQDRDVLVLOG 616  
DB 599 GRFPNEKQLKQLOGLMHTY-----FPNKGTQOYTDQIE-RPLMWGSVMNRRAHYES 651  
QY 617 PIWAKIPTHDEHFPs-PLMGFGILKNPPQILIKNTPVPAHPAEFSATKFASFITQYS 675  
DB 652 QLMSKIPLNDDSEFKTQFAALGSGWGLHQPPQILFK--ILPQSGPIGIGIKSMGITTLVQYA 709  
QY 676 TGQVSVEIEMEL-QKENSKRWNPE 698  
DB 710 VGIWTVMTFKLGPRKATGRWNPQ 733

RESULT 4  
COAL\_PAVP9  
ID COAL\_PAVP9 STANDARD; PRT; 729 AA.  
AC P33484;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Coat protein VP1 [Contains: Coat protein VP2].  
OS Porcine parvovirus (strain 90HS) (ppv).  
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.

OX NCBI\_TaxID=33725;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89319168; PubMed=2750278;  
RA Sakurai M., Nishimori T., Ushimi C., Nakajima H.;  
RT "Nucleotide sequence of capsid protein gene of porcine parvovirus.";  
RL Virus Res. 13:79-86(1989).  
CC -1- SUBUNIT: MATURE VIRION CONTAINS THREE CAPSID PROTEINS DESIGNATED  
CC VP1, VP2, AND VP3 AND A NONCAPSID PROTEIN NS-1.  
CC -1- MISCELLANEOUS: VP3 MIGHT BE A POSTTRANSLATIONAL CLEAVAGE PRODUCT  
CC OF VP2 IN SEVERAL AUTONOMOUS PARVOVIRUSES.  
CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.  
DR HSSP; P30129; 4DPV.  
DR InterPro; IPR001403; Parvo\_coat.  
DR Pfam; PF00740; Parvo\_coat; 1.  
KW Coat protein; Glycoprotein.  
FT CHAIN 1 729 COAT PROTEIN VP1.  
FT FT 151 729 COAT PROTEIN VP2.  
FT CARBOHYD 172 172 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 198 198 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 282 282 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 433 433 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 471 471 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 573 573 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 604 604 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 651 651 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 729 AA; 80938 MW; 92538BFF9A0C78E6 CRC64;

Query Match 12.1%; Score 482; DB 1; Length 729;  
Best Local Similarity 25.7%; Pred. No. 1.1e-22;  
Matches 197; Conservative 112; Mismatches 314; Indels 144; Gaps 35;

QY 43 RGLVLPGYKYLGPFGNGLDKGEPVNAADAALAEHDKAYDQOLKAGDNPYLRYNHADAEFOE 102  
DB 9 RGLTLPGYKYLGPFGNSLDQGEPTNPSDAAAKEHDEAYDKYIKSGKNPYFYFSADEKFIK 68  
QY 103 RLQEDTSFGNLAGRAVFOAKKRVLEPLGLVEEAKTAPGKKRPVEQSPQEPDSSSGIGK- 161  
DB 69 ETEHAKDYGKIGHYFFRA-KRAFRP-KLSETDSPTTSQOPEVRRSRPRKHPGSKPRKR 125  
QY 162 -----TGQOPAKKRLNFGOTGDSSEVPDPPLGEPBPATPAVGPPTMASGGAPMA 212  
DB 126 PAPRHIFINLAKKAKGTSTNNSMSSENVEQHPINAGTEL-SATGNESGGGGGG--- 181  
QY 213 DNEGADGVGNASGNWH--CDSTWLGD--RVITSTRTWALPTYYNNHLKYQIS--SAST 265  
DB 182 -GGRGAGGVGVSTGSFNNQTEFOYLGELVRITAHASRLIHLNMPHEHTYKRIHVAINSES 240  
QY 266 GAS----NDNHFGYSTPWGYFDENRHFCHFSRPDMQRLINNNGFRPKRLNFKLFNIOV 321  
DB 241 GYAGQWQDDAHQTQWTPWSLIDANAMGVWFNPADWQLISNMTEINLVSFEQEIFNVVL 300  
QY 322 KEVT--TNDGVTTIANNLSTVQVFSDESEYQLPYVLGSAHQGLPFPADVFMIPOGY 378  
DB 301 KTLTESATSPPTKIYNNDLTASLWALDTNNTLPTPAAPRSETLGFYPMWPTKPTQYRY 360  
QY 379 L-----TLNNGSQAVERS-----SFYCLE-YFPSOMLRTGNNF-TFSYTFE 417  
DB 361 YLSTCRNLNPPTYTGSSQQTIDSIQTGLHSDIMFYTIENAVPIHLRLTGDEFSTGIYHFD 420  
QY 418 EVPFHSSYAHQSOLDR-----LMNPLI--DQYLYLNRTQNGSGAQNKDLLFSRGS 467  
DB 421 TKPL--KLTHSWQTNRSLLGPLKLLTPTTEGDQHPGTLPAANTRKGYHQTMNNSYTEAT 478  
QY 468 PAGMSVQPKNMLPGPCYRQQRVSKTKTDNNNSNFTWTGASKYNLNGRESIINP--GTAMA 525  
DB 479 ---AIRP-----AQVGYNTPYMNFESYNGGPF-----LTPIVPTADT 512  
QY 526 SHKDEDEKFFPMSCVMIFGKESAGASNTALDNVMITDEEIKATNP-----VATERG 579  
DB 513 QYNDDE----PNGAIRFTWGYQHGLTTS-----SQELERYTFNPQSKGRAPKQGFNO 562

[illegible]

```

RESULT 5
COAL_PAVPN
ID      COAL_PAVPN      STANDARD;      PRT;      729 AA.
AC      P18546; Q89816;
DT      01-NOV-1990 (Rel. 16, Created)
DT      01-NOV-1990 (Rel. 16, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Coat protein VP1 [Contains: Coat protein VP2] (Version 1).
OS      Porcine parvovirus (strain NADL-2) (PPV).
OC      Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
OX      NCBI_TaxID=10797;

```

RP SEQUENCE FROM N.A.  
RX MEDLINE=91021005; PubMed=2219713;  
RA Vasudevacharya J., Basak S., Srinivas R.V., Compans R.W.;  
RT "The complete nucleotide sequence of an infectious clone of porcine  
RL parvovirus, strain NADL-2.";  
Virology 178:611-616 (1990).

RP SEQUENCE FROM N.A.  
MEDLINE=90085785; PubMed=2596019;  
RA Vasudevacharya J., Basak S., Srinivas R.V., Compans R.W.;  
RT "Nucleotide sequence analysis of the capsid genes and the right-hand  
RL terminal palindrome of porcine parvovirus, strain MADL-2.",  
Virology 173:368-377(1989).

CC -1- SUBUNIT: MATURE VIRION CONTAINS THREE CAPSID PROTEINS DESIGNATED VP1, VP2, AND VP3 AND A NONCAPSID PROTEIN NS-1.

CC -1- MISCELLANEOUS: VP3 MIGHT BE A POSTTRANSLATIONAL CLEAVAGE PRODUCT OF VP2 IN SEVERAL AUTONOMOUS PARVOVIRUSES.

CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collabora  
CC between the Swiss Institute of Bioinformatics and the EMBL outsta  
CC the European Bioinformatics Institute. There are no restrictions o  
CC use by non-profit institutions as long as its content is in n  
CC modified and this statement is not removed. Usage by and for comm

CC entities requires a license agreement (See <http://www.isb-sib.ch/ann>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC	EMBL: M38367; AAAA46919.1; -
DR	EMBL: M38367; AAAA46921.1; -
DR	EMBL: M32787; AAAA46917.1; -
DR	EMBL: M32787; AAAA46918.1; -
DR	PIR: B33743; VCPVNA.
DR	UCCD: D20120; 4DPRV.

DR HSSP; F50125; 4DEV.  
DR InterPro; IPR001403; Parvo\_coat.  
DR Pfam; PF00740; Parvo\_coat; 1.  
KW Coat protein; Glycoprotein.

FT	CHAIN		COAT PROTEIN VP1.	
FT	1	729	COAT PROTEIN VP2.	
FT	151	729	N-LINKED (GLYCINE)	1 (POTENTIAL)
FT	173	172		

	FT	CARBONHYD	CARBONHYD	FTT
198	172	198	282	282
N-LINKED	(GLCNAC . .)	(GLCNAC . .)	N-LINKED	(GLCNAC . .)
POTENTIAL	.	.	(POTENTIAL)	.
POTENTIAL	.	.	(POTENTIAL)	.

FT	CARBOHYD	330	N-LINKED (GLCNAC . . .)	(POTENTIAL) .
FT	CARBOHYD	433	N-LINKED (GLCNAC . . .)	(POTENTIAL) .
FT	CARBOHYD	471	N-LINKED (GLCNAC . . .)	(POTENTIAL) .
FT	CARBOHYD	433	N-LINKED (GLCNAC . . .)	(POTENTIAL) .

FT	CARBOHYD	573	573	N-LINKED (GLCNAC. . .)	(POTENTIAL) .
FT	CARBOHYD	604	604	N-LINKED (GLCNAC. . .)	(POTENTIAL) .
Em	CARBOHYD	661	661	N-LINKED (GLCNAC. . .)	(POTENTIAL) .

FT	DOMAIN	175	190	GLY-RICH.
SO	SEQUENCE	729 AA;	80946 MW;	EF816E246C80DB42 CRC64;
Query Match		12.1%;	Score 482;	DB 1; Length 729;
Best Local Similarity		25.5%;	Pred. NO. 1.1e-22;	
Matches 197;	Conservative	111;	Mismatches 310;	Indels 154; Gaps 35;

Qy	43	RGLVLP	PGYKYL	GFPNGL	DKEGP	VNAA	DAAL	LEHDK	AYDQ	LKAG	DNPLY	RYNH	ADAE	FQOE	102
Db	9	RGLTLP	PGYKYL	GFPGNS	LDQGE	PPTNP	SDAA	KEHDE	AYDKY	IKSG	KNPY	FYFSA	DADE	KFIK	68

```

QY      103 RLQEDTSGFGNIGRAVEQAKRRVLEPLGIVEGAKTAPGKRRPVEQSPQEDSSSGIGK- 161
      :  ||  :  ||  :  ||  :  ||  :  ||  :  ||  :  ||  :  ||  :  ||  :  ||
Db      69 ETEHAKDYGGKIGHYFFRAKRAFAPKL---SETDPTTSQQPEVRRSRPKRPGSKPPGKR 125

```

**D**b            102         :| | | :| | :| |  
            106 PAPHIFINLAKKAGTSNTNSMSENEQHPINAGTEL-SATGNESGGGCGGC--- 181

182 -GGRGAGGVTGTENNQTEFQYLGEGLVYRITAAASRLIHLNMPHEHETFKRIHVLNSES 240

Db	241	GVAGQWQDDAHTQWTPWLSLIDANAGCWAFNPADWQOLISNNMTEINLVSEQEI	300
Oy	322	KEVT---TNDGVTIANNLSTVQVFSDEYOLPYVLGSAHQGLPPFADVEMIPQGY	378

Db	301	KTITESATSPPTKIYNNDLTASLWALDTNNLTLPYTPAPRSETLGFYPMLEPKPTQYRY	360
QY	379	L-----TLNGSQAVGRS-----SFYCLE-YFPSQMLRTGNNE-TFSYTFE	417

Db 361 YLSCIRNLNPYYTGQSQITDSIQGTGHSIDIMFYTIENAVPIHLRTGDEFSTGIYHFD 420

Qy 418 EVPFHSSIAHSQSLDR-----LNPLI--DQLYYLNRTQNGSGAONKDLLFSRGS 467

```

DO      421 INFL--KULISWQJNKSUGSUFENLIEFI1EGDQJFGI1EFAMININGINVA1INNSII1EAI 7/0
QY      468 PAGMSVQPKMWLBPFCYRQQRVSKTKTDNNNSNFTWTGASKYNLNGRESIINP--GTAMA 525
          ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

526 SHKDEDEKFFPMSCVMI FGKESAGASNTALD--NMITDEEIK--ATNP-----VAT 574

QY	575	ERFGTVA-VNFQSSS-----TDPATG--DVHMGALP-----GMVQDRDYYLQGPIMA	620
Dh	558	OOFNQOAPLNTENTNNGTLPSDPIGKSNVHEMNTLNTYGPLTALNNTAEPVEPNGOIWD	617

```

QY      621 KIPHTD--GHFHPSPLMGGFGLK-NPPQQLIKNTVPANPRAEFSA-TKEASFITQYST 676
      |  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      618 KELDTDLKPRH---VTAPFVCQKNPPGQLEVKIAP---NLTDDEFNADSPQOPRIITYSN 671

```

```

QY      677 GQSV E I E M E L O K E N S K R W N P E V Q Y T S N A K S A N V D F T V D N N G L Y T E R P I G 728
      : : : : | | | | : : : : | : | | | |
Db      672 F W W K G T L T T A K M R S S N M W N P I Q O H T T ----- T A E N I G N Y I - P T N I G 712

```

## RESULT 6

### COAL | PAYPK

ID	TC01 PAVPK	STANDARD;	PRT;	729 AA.
AC	P52501;			
DT	01-OCT-1996 (Del	34	Created)	

DT	Rel.	last annotation update)
01-OCT-1996	Rel. 34,	last sequence update)
16-OCT-2001	Rel. 40,	last annotation update)

DE Coat protein VP1 (contains: Coat protein VP2) .  
OS porcine parvovirus (strain Kresse) (PPV) .  
OC Viruses: ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus .

OX	NCBI_TaxID=73487;
RN	[1]
RD	SEQUENCE FROM N









RL J. Gen. Virol. 71:2747-2753(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CU-4;  
RX MEDLINE=91272479; PubMed=1647068;  
RA Parrish C.R.;  
RT "Mapping specific functions in the capsid structure of canine  
RT parvovirus and feline panleukopenia virus using infectious plasmid  
RT clones.";  
RL Virology 183:195-205(1991).  
CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X55115; CAA38911.1; -  
DR EMBL; M38246; AAC37928.1; -  
DR EMBL; M38246; AAC37929.1; -  
DR PIR; B36608; VCPVFP.  
DR HSSP; P30129; ADPV.  
DR InterPro; IPR001403; Parvo\_coat.  
DR Pfam; PF00740; Parvo\_coat\_1.  
KW Coat protein; Glycoprotein.  
FT CHAIN 1 727 COAT PROTEIN VP1.  
FT CHAIN 144 727 COAT PROTEIN VP2.  
FT DOMAIN 165 190 GLY-RICH.  
FT CARBOHYD 168 168 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 190 190 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 207 207 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 323 323 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 586 586 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 648 648 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 660 660 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT DOMAIN 165 180 GLY-RICH.  
SQ SEQUENCE 727 AA; 80386 MW; 648596C09B621FF5 CRC64;  
Query Match 10.8%; Score 431.5; DB 1; Length 727;  
Best Local Similarity 23.5%; Pred. No. 1.5e-19;  
Matches 185; Conservative 117; Mismatches 297; Indels 189; Gaps 35;  
QY 43 RGLVLPGYKYLGPENGLDKGEFVNADAALEHDKAYDQQLKAGDNFYLRYNHADAERQ 102  
DB 10 RGLVPPGYKYLGPENGLDQGEPTNPSDAAKEHDEAYAYLRSGKNPYLFSPADQRFID 69  
QY 103 RLQEDTSPFGNIGRAVFOAKKRVLEPLGLVEEAGTAPGKRPRVEQSPQEDSSGIGKT 162  
DB 70 CTKDAKDWGKXIGHYFFRAKKAIAPLVLTDPHSTSRPTK.PTKRSKPPHIFINLAK- 127  
QY 163 GQPAKKRLNFGQTSSESVP---DPQPLGEPPA--TPAVGPTTMASSGGGAPMADNNE 216  
DB 128 -----KKKAGAGQVRDNLAPMSDGAQVDPDGGQPAVRNERATSGSGSGGG-----GG 176  
QY 217 GADGVGNASGNMH-----CDSTWLGDRTVITSTRTALPTYNHLYKQI-----SS 262  
DB 177 GSGGVGISTGTFFNNQTEFKFLENGWV--EITANSRLVHLNMPESSENYKRVVYVNMMDKTA 234  
QY 263 ASTGASNDNHFGYSTPWGYFDNRFCHFSRPDQRLINNNWGFRPKRLNFKLFIQVK 322  
DB 235 VKGNMALDDIHQIVTPWSLVDANAWGVFENPGDWQLIVNTISELHLVSEQEIENVLK 294  
QY 323 EVT---TNDGVTTIANNLSTVQVFSDESEYQLPYVLGSAHQ3CL-----PPFPA----- 368  
DB 295 TVSESATQPTKYVNNDLTASIMVALDSNNTMPTPAAMRSITLGFYPMKRTIPTPMRY 354  
QY 369 ---DVFMIPOYGYLTLNNGSAVGRS-----SFYCL3-YFPSQMLRTGNNE-TFS 413  
DB 355 FQWDRTLIPSH-----TGTSGTPTNVHGTDPDDVQFYTIENSVPVHLRTGDEFATGT 408

QY 414 YTFEEVPHSSVYAHQSOLDRLMN--PLIDQLYLNRITQNGSGSAQNKDLFSGSPAGM 471  
DB 409 FFFDCKP--CRLTHTWQTRNRLGLPP-----FLNSLPQSEGATNFGDIGVQDKRRGV 459  
QY 472 S-----VQPKN-WLPGCYRQQRVSK-----TKTDNNNSNFTWT 504  
DB 460 TOMGNITDITEATIMRPAEYGVSAFYSEASTQGFPKTPIAAGRGGAQTDENOAA---D 516  
QY 505 GASKY---NLNGRESIINPGT-----AMASHKDEDEKFFPMGVMIFGESAGASNTALD- 556  
DB 517 GDRPYAFGRHQGQKTTTGETPERFTYIAHQT-----GRYPEGDWIQINIF 563  
QY 557 NMVITDEEIKATNPVATERFGYVAVNFQSSSTDPATGDVHAMGALLPGMWQDRDYVLOG 616  
DB 564 NLPVTNDNVLLPTDPDG---GKTGINY--TNIFNTYGPLTALNNVF-----PVPYNG 610  
QY 617 PIWAKIPHTDGHFHPSPLMGFGIK-----NPPQILIKNTVPVAN---PPA 660  
DB 611 QIWDKEPDT-----LKPRLHVNAFPVCCQNNCPGOLFVKVAPNLTNEYDPDA 657  
QY 661 EFSATKPSFITQYSTGVSEIEWELQENSKRWNPVEQYTSNYAKSANVDFTVDNGL 720  
DB 658 SANMSR-----IVTYSDFWKKGLYFKAKLRASHTWNPICQMSIN-----VDNQFN 703  
QY 721 YTEPRPIG 728  
DB 704 YV-PNNIG 710

RESULT 11  
COAT\_PAVCB STANDARD; PRT; 727 AA.  
ID COAT\_PAVCB

AC 011213;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Coat protein VP1 [Contains: Coat protein VP2].  
OS Canine parvovirus (strain B) (CPV).  
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.  
OX NCBI\_Taxid=59284;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91272479; PubMed=1647068;  
RA Parrish C.R.;  
RT "Mapping specific functions in the capsid structure of canine  
RT parvovirus and feline panleukopenia virus using infectious plasmid  
RT clones.";  
RL Virology 183:195-205(1991).  
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHEDRAL UNITS,  
CC OF A COMBINATION OF VP2, VP3, AND SOME VP1.  
CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M38245; AAB02799.1; -  
DR EMBL; M38245; AAB02800.1; -  
DR HSSP; P30129; ADPV.  
DR InterPro; IPR001403; Parvo\_coat.  
DR Pfam; PF00740; Parvo\_coat\_1.  
KW Coat protein; Glycoprotein.  
FT CHAIN 1 727 COAT PROTEIN VP1.  
FT CHAIN 144 727 COAT PROTEIN VP2.  
FT CARBOHYD 168 168 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 190 190 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 207 207 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 323 323 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 586 586 N-LINKED (GLCNAC. . .) (POTENTIAL).



FT CARBOHYD 648 648 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 660 660 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 727 AA; 80342 MW; 115F3E1A79098EBE CRC64;

Query Match 10.8%; Score 430.5; DB 1; Length 727;  
Best Local Similarity 23.2%; Pred. No. 1.8e-19;  
Matches 184; Conservative 112; Mismatches 297; Indels 201; Gaps 32;

QY 43 RGLVLPYGYKYLGPENGLDKGEVNAADAALHDKAYDQOLKAGDNPYLRYNHADAFFQOE 102  
DB 10 RGLVPPGYKYLGPENGLDQGEPTNPSDAAKEHDEAYAYLRSGKNPYLYFSPADQRFID 69  
QY 103 RLQEDTSFGNLRGAVFQAKKRVLEPLGLVEGAKTAPGKKRPVEQSPQEPDSSGIGKT 162  
DB 70 QTKDAKDWGKIGHYFFRAKKAIAIPVLTPTDHPSTSRPTK-PTKRSKPPHIFINLAK- 127  
QY 163 GQOPAKKRLNFGQTDSESV---DPQPLGEPRA--TPAAVGPTMASGGGAPMADNNE 216  
DB 128 ----KKKAGAGQVKRDNLAPMSDGAQVQPGQPAVRNERATSGNGSGGG-----GG 176  
QY 217 GADGVGNASGNWH-----CDSTWLGDRVITSTRTVALPTYNHLYKOI-----SS 262  
DB 177 GSGGVGISTGFENNQTEFKEFLENGWV--EITANSSRLVHLNMPSESNYRVRVNNMDKTA 234  
QY 263 ASTGASNDNHYFGYSTPWGYFDNFRFHCHSPRDMQRLINNNGFRPRKLNFKLFNIQVK 322  
DB 235 VNGNMALDLDIAQIVTPWSLVDANAMGVWENPGDWQLIVNTMSELHVSFEQEIFNVVLK 294  
QY 323 EVT---TNDGVTIANNLSTVOVFSDEYQLPVYLSAHQGL-----PPFPA----- 368  
DB 295 TVSESATQPTPKVYNNDLTASLWALDSNNMTPTPAMRSETLGFYPMKPTIPTPWRYX 354  
QY 369 ---DVFMIPOYGYLTLNNGSQAVGRS-----SFYCLE-YFESQMLRTGNF-TFS 413  
DB 355 FQWDRITLPSH-----TGTSGTPTNIYHGTDPDDVQFYTIENSVPVHLRLTGDEFAGT 408  
QY 414 YTFEEVFPFSSYASHQSOLDRLMN--PLIDQYLYLNRTQNGSGAONKDLFSRGPAGM 471  
DB 409 FFFDCKP--CRLTHWTQNRALGLPP-----FLNSLPQSEGATNFGDI-----GV 451  
QY 472 SVQPKNWLPGPCYRQQRVSKTKTDNNNSNF-----TWTGASKYNLNGRESIINP- 520  
DB 452 -----QODKRGVTOQMGNTNYITEATIMRPAVGVSAFYYSFEASTQGPFX 497  
QY 521 ---GTAMASHKDEDEKFFPMGSGVMIFFKESAGASNTALD----- 556  
DB 498 TPIAAGRGAGQTDENQADGNPRYAFGRQHQKTTTGTGETPERFYIAHQDTGRYPGDM 557  
QY 557 -----NVMITDEEIKATNPVATERFRTYAVNVFOSSSTDPATGVHAMGALPGMWQDR 610  
DB 558 IONINENLPTNDNVLLPTDPIG---GKTGINY--TNIFNTYGPLTALNNVP----- 604  
QY 611 DVLQGPITWAKIPHTDGHFHPSPLMGFGIK-----NPPQILIKNTPVPPAN 657  
DB 605 PVPYPRNGQIWDKEFDTD-----LKPRLHVNAPFVQCNNGCPGLFVKVAPNLTN 651  
QY 658 ---PPAESATKFAFITQYSTGVSVETIEMELQKENSKRWNPEVQYTSNYAKSANVDT 714  
DB 652 EYDPDASANSR---IVTYSDFWVGKLVFKAKLRASHWTNPIQOMSIN----- 697  
QY 715 VDNNGLYTEPRPIG 728  
DB 698 VDNQFNYY-PSNIG 710

RESULT 12  
COAT MUMIM STANDARD; PRT; 718 AA.  
AC P07302; Q9WMH3; Q9WMH2;  
DT 01-APR-1988 (Rel. 07, Created)  
DT 01-APR-1988 (Rel. 07, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Coat protein VP1 [Contains: Coat protein VP2].

OS Murine minute virus (strain MWMI) (Murine parvovirus).  
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.  
OX NCBI\_TaxID=10795;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=86115415; PubMed=3502703;  
RA Astell C.R., Gardiner B.M., Tattersall P.;  
RT "DNA sequence of the lymphotropic variant of minute virus of mice, MVM(i), and comparison with the DNA sequence of the fibrotropic prototype strain.";  
RL J. Virol. 57:656-669(1986).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=85242059; PubMed=3855242;  
RA Sahli R., McMaster G.K., Hirt B.;  
RT "DNA sequence comparison between two tissue-specific variants of the autonomous parvovirus, minute virus of mice.";  
RL Nucleic Acids Res. 13:3617-3633(1985).  
RN [3]  
RP X-RAY CRYSTALLOGRAPHY (3.5 ANGSTROMS) OF 132-718.  
RA Iltamas-Salz A.L., Agbandje-Mckenna M., Winkler W.R., Bratton J.,  
RA Tattersall P., Rossmann M.G.;  
RT "Structure determination of Minute Virus of mice.";  
RL Acta Crystallogr. D 53:93-100(1997).  
CC -!- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; X02481; CAB46507.1; -.  
DR EMBL; X02481; CAB46508.1; -.  
DR EMBL; M12032; AAA69569.1; ALT\_INIT.  
DR PIR; B23008; VCPVIM.  
DR PDB; 1MVM; 25-FEB-98.  
DR InterPro; IPR001403; Parvo coat.  
DR Pfam; PF00740; Parvo coat; 1.  
KW Coat protein; Glycoprotein; 3D-structure.  
FT CHAIN 1 718 COAT PROTEIN VP1.  
FT CHAIN 132 718 COAT PROTEIN VP2.  
FT CARBOHYD 180 180 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 219 219 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 502 502 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 635 635 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 671 671 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT DOMAIN 157 172 GLY-RICH.  
FT CONFLICT 144 144 A -> G (IN REF. 2).  
SQ SEQUENCE 718 AA; 79017 MW; B43C1762ED5F74B9 CRC64;

Query Match 10.6%; Score 424.5; DB 1; Length 718;  
Best Local Similarity 23.7%; Pred. No. 4.1e-19;  
Matches 184; Conservative 104; Mismatches 298; Indels 189; Gaps 31;

QY 45 LVLPYGYKYLGPENGLDKGEVNAADAALHDKAYDQOLKAGDNPYLRYNHADAFFQERL 104  
DB 1 MVPPGYKYLGPENGLDQGEPTNPSDAAKEHDEAYDQYIKSGKNPYLYFSAADQRFIDQ 60  
QY 105 QEDTSFGNLRGAVFQAKKRVLEPLGLVEGAKTAPGKKRPVEQSPQEPDSSGIGKTG- 163  
DB 61 KDAKDWGKIGHYFFRTKRAFAPKLATDSE-----FG-----TSGVSRAGK 101  
QY 164 -----QOPAKKRLNFG---QTDSESVDPQPLG-----EPATPAVGPPTM 203  
DB 102 RTRPPAYIFINQARAKKLTSSAAQSSQTMDSGTSQPDGNAVHSAARVERADGPG-- 159  
QY 204 ASGGAPMADNNEGADVGNASGNMHCDS--TWLGD--RVITSTRTVALPTYNHLYK 258  
DB 160 GSGGG-----GSGGGGVSTGSDYDQTHYRLDGVWEITATLRLVHLNMPKSENYC 213

```

QY 259 QI-----SSASTGASNDNHFGYSTPWGYFDENRFHCHFSPRDMORLINNNGFRPKR 311
Db 214 RIRVHNTTDTSVKGNMAKDADAHEQIWTWPSLVDANAMGWLQPSDMQYICNTMSQNLVS 273
QY 312 LNFKLENIQVKEVTND---GVTTIANLSTVOVFSDEYOLPYVLGSAHQGLPPF 367
Db 274 LDQELFNVLKTVTEQDSGGQAIKYNNDLTACMVAVDSNNILPYTPAANSMTLLGFYP 333
QY 368 ADVEMIPQYGY-----LTLNN-----GSQAVGRSSFYCLEYFBS-QMLRTG 407
Db 334 WKPTIASPYRYFFCVDRLDLSVTYENQEGTIEHNWGTGPKMNSQFTIENTQOITLLRTG 393
QY 408 NNF-TFSYTFEEVFPFHSSYASHQSOLDRLNN--PLIDQYLYYIANTONQSGSAQNKDLIFS 464
Db 394 DEFATGTYYFDTNPV--KLTHTWQTNRLQGPBLLSTF---PEADTDAGT-----LTA 441
QY 465 RGSPPAGMSVQPKNWL-----PGPCYRQQRVSKTKTDNNNSN 500
Db 442 QGSRHGATQMEVNWVSEAIRTRPAQVGFQCPHNDFEASRAGP-FAAPKVPADVTQGVDR 500
QY 501 FTWTGASKYNLNGRESIINPGTAMASHKDEDEKF-----FPMGVMIFGESAGAS 551
Db 501 ANGSRRYSYKQHGEMWAHGPAPERYTWEDENFGSGRDRIGFIQSAPLVVPPLNGIL 560
QY 552 NTA LDNVMITDEEIKATNPVATERFGTVAVNEQS--SSTDIPATGDVHAMGALPGWVQD 609
Db 561 TNA-----NPIGTKN---DIHFSNVFNSXGPLTAFSH-----P 590
QY 610 RDVYLOGPIWAK---IPHTDGHFHPSPLMGFGFLKNPPOILIKNTP--VPANPBAEFS 664
Db 591 SPVYPOQOIMDKELDLEHKPRLHITAPFV---CKNNAPOQMLVRLGPNLTDQYBENGATL 647
QY 665 TKPASFITQYSTGOVSVEIEMELQKENSKRWNPEVOQTSNYAKSANDFTVDNNG 719
Db 648 SRIVTYGTFFWKGLTMRAKLRA---NTTWNPYQO-----VSVEDNG 686

```

```

RESULT 13
COAT_PAVCN
ID COAT_PAVCN STANDARD; PRT; 748 AA.
AC P12930; Q84387;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Coat protein VP1 [Contains: Coat protein VP2].
OS Canine parvovirus (strain N) (CPV).
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
OX NCBI_TaxID=10791;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88062992; PubMed=2824850;
RA Reed A.P., Jones E.V., Miller T.J.;
RT "Nucleotide sequence and genome organization of canine parvovirus.";
RL J. Virol. 62:266-276(1988).
CC -1 SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,
OF A COMBINATION OF VP2, VP3, AND SOME VP1.
CC -1 SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M19296; AAA67460.1; -.
DR EMBL; M19296; AAA67461.1; -.
DR PIR; B29962; VCPVCP.
DR HSSP; P30129; 4DPV.
DR InterPro; IPR001403; Parvo_coat.
DR Pfam; PF00740; Parvo_coat; 1.
KW Coat protein; Glycoprotein.

```

FT	CHAIN	1	748	COAT PROTEIN VP1.
FT	CHAIN	165	748	COAT PROTEIN VP2.
FT	CARBOHYD	189	189	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	211	211	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	228	228	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	344	344	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	607	607	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	669	669	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	681	681	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	DOMAIN	186	203	GLY-RICH.
FT	CONFLICT	27	27	M -> XLSM (IN REF. 1; AAA67460).
SO	SEQUENCE	748 AA;	82715 MW;	13E487C6862C73E6 CRC64;

Query Match	10.6%;	Score 421.5;	DB 1;	Length 748;
Best Local Similarity	23.0%;	Pred. No. 6.7e-19;		
Matches 184;	Conservative 114;	Mismatches 287;	Indels 215;	Gaps 34;

QY	44	GLVLPGYKYLGPFGNLGKGPVNAADAALAEHDKAYDQOLKAGDNPLYRYNHADAEPQER	103
Db	32	GLVPPGYKYLGPFGNSLDQGEPTNPSSDAAKEHDEAAYLRSGKNPYLYFSPADQRFIDQ	91
QY	104	LQEDTSFGNLRGAVFOAKKRVLEPLGLVEGAKTAPGKKRPVEQSPQEPDSSSGIGKTG	163
Db	92	TKDAKDWGKIGHYFFRAKKAIAFVLTDTPDHPSTSRPTK-PTKRSKPPHIFINLAK--	148
QY	164	QQPAKRLNFGQTDSESVP---DQPLGEPPA--TPAAVGPTTASGGGAPMADNNEG	217
Db	149	----KKKAGAGQVKRDNLAPMSDGAVQPDGGQPAVRNERATSGSGSGGG-----GGG	198
QY	218	ADGVGNASGNMH-----CDSTWLGDRVITSTRTRWALPTYNNHLYKQI-----SSA	263
Db	199	SGGVGISTGTFFNNQTEKFLNGWV--EITANSSRLVHLNMPESERNYRRVVVNNMDKTAV	256
QY	264	STGASNDNHFGYSTPWCYFDENRFHCHFSPRDMORLINNNWGPFRKLNFKLFNIQVKE	323
Db	257	NGNMALDDIHAQIVTPMSLYDANAMGWFMNPGDMQOLIVNTMSEHLVSFEQEIFNVVLKT	316
QY	324	VT---TNDGVTTIANNLTSVQVFSDEYQLPYVLGSAHQGL-----PPFA-----	368
Db	317	VSESATQPTKVVYNNDLTASIMVALDSNNTMPTPAAMRSETLGFYFWKPTIPTPWRYE	376
QY	369	--DVFMIPQYGYLTLNNGSQAVGRS-----SFYCLE-YFPSQMLRTGNNF-TFSY	414
Db	377	QMDRTLIPSH-----TGSCTPTNIYHGTDPDDVQFYTIENSVEVHLRTGDEFATGTF	430
QY	415	TPEEVPFHSSYAHQSQSLDRLMN--PLIDQYLYLNRTQNGSGSAQNKDLLFSRGPAGMS	472
Db	431	FFDCKP--CRLTHTWQTNRALGLP-----FLNSLPQSEGATNPGDI-----GV-	472
QY	473	VQPKNWLPGPCYRQQRVSKTKTDNNNSNF-----TWGASKYNLNGRESIINP-GT	522
Db	473	-----QODKRGVTOGMGNTNYITEATIMRPAEVGYSAPYYSFEASTQGPFKT	519
QY	523	AMASHK-----DDEKKEFFPMGCVMI FGKESAGASNTALD-----	556
Db	520	PIAAGRGGAQTYENQADGPPRY-----AFGRQHGQKTTTGETPERFTYIAHQDTGR	572
QY	557	-----NVMITDEEIKATNPVATEREGTVAVNPQSSSTDTPATGDVHAMGALPG	604
Db	573	YPEGDWIQNINENLPTVNDNVLLPTDPIG---GKTGINY--TNIFNTYGPALTANNVP-	625
QY	605	MWVQDRDVYLLQGPWAKIPHTDGHFHPSPMLGGFGIK-----NPPQILIKN	651
Db	626	-----PVPNGQIWDKEPFD-----LKPRLHVNAPEVCQNNCPGQLFVKV	666
QY	652	TPVPAN---PPAEFSATKFAFITQYSTGQVSVEIEMELQKENSKRWNPEVQYTSNYAKS	708
Db	667	APNLITNEYDPDASANMSR---IVTYSDFWKKGLVFKAKLRAHSHTWNPICQMSIN----	718
QY	709	ANVDFTVDNNGLYTEPRPIG 728	
Db	719	-----VDNQFNVY-PSNIG 731	

```

RESULT 14
COAT_PAVCD
ID COAT_PAVCD STANDARD; PRT; 737 AA.
AC P17455;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Coat protein VP1 [Contains: Coat protein VP2].
OS Canine parvovirus (strain CPV-D Cornell 320) (CPV).
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirus.
OX NCBI_TaxID=10790;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89020796; PubMed=3176341;
RA Parish C.R., Aguadro C.F., Carmichael L.E.;
RT "Canine host range and a specific epitope map along with variant
RT sequences in the capsid protein gene of canine parvovirus and related
RT feline, mink, and raccoon parvoviruses.";
RL Virology 166:293-307(1988).
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHEDRAL UNITS,
CC OF A COMBINATION OF VP2, VP3, AND SOME VP1.
CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

[illegible]

Query Match	10.2%;	Score 407.5;	DB 1;	Length 737;
Best Local Similarity	22.8%;	Pred. No. 4.9e-18;		
Matches 181;	Conservative 112;	Mismatches 301;	Indels 201;	Gaps 32;

```

QY      42 GRGLVLPGYKYLGPFPNGLDKGEFPVNAADAALLHDKAYDQOLKAGDNPYLRYNHADAFFQ 101
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      19 GEGKDLITYKYLGPNGSLDQGEPTNPSDAAAKEHDEAYAAYLRSKGKNPYLXSPADQRFI 78

QY      102 ERLQEDTSFGGNLGRAVFOAKKRVLLEPLGVEBGAKTAPGKKRPVEQSPQEPDSSGIGK 161
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      79 DQTKDAKDWGSKI GHYFFRAKKAIAIAPVLTDPDHPSTSRPTK - PTKRSKPPHIFINLAK 137

QY      162 TGQQPAKKRLNFGQTGDSSEVP --- DPQPLGEPPA - -TPAVGPTTWAAGGAPMADNN 215
      | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      138 -----KKKAGAGQVKRDNLAPMSDGAVOQPDGQPAVRNERATGSGNGSGGG-----G 185

QY      216 EGADGVGNASGNWH-----CDSTWLGDREVITSTRTWALPTYNNHLKYQI-----S 261
      | : | | | : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      186 GSGSGVGISTGTFFNNQTEFKFLENGWV - -ETANSSRLVHLNMPESSENYRRVVVNNMDKT 243

QY      262 SASTGASNDNHYFGYSTPMWGYFDENRFCHSPRDPQRLINNNGFRPKRLNFKLFNIQV 321
      : : : : : : : : | | | | | | | | | | | | | | | | | | | | | | : : : : :
Db      244 AVNGENNALDDIIHAQIVTPWSLV DANAWGVENEGDWQOLIVNTMSELHLVSEFOELFNVVL 303

```

QY	322	KEVT--TNDGVTTIANNLSTVQVFSDEYQLPYVLGSAHQGL-----PPFPA----	368
Db	304	KTVSESATQPTKVYNNDLTASLMVALDSNNTMPTPAMRSETLGFYWKPTIPTPWRY	363
QY	369	----DVFMIPQYGYLLTLNNGSAVGRS-----SFYCLE-YFPQMLRTGNNE-TF	412
Db	364	YFQWDRTLIPSH-----TGTSGPTNIYHGTDPPDVQFYTIENSVPYHLRTGDEFATG	417
QY	413	SYTFEEVPFHSSYAHQSGLDRLMN--PLIDQYLYLNRTONQSSGAQNKDLFSRGPAG	470
Db	418	TFFFDCKP--CRLTHTWQTNRALGLPP-----FLNSLPQSEGATNFGDI-----G	460
QY	471	MSVQPKNMLPGPCYRQQRVSKTKTDNNSNF-----TWGASKYMLNGRESIINP-	520
Db	461	V-----QQDKRGVLTQMANTNYITEATIMRPAEVGYGAPYSPFASTQGP	506
QY	521	----GTAMASHKDEDEKFFPMSGVMI FGKESAGASNTALD-----	556
Db	507	KTPIAGRGAQOTDENQAADGNRYAFGRQHQOKTTTGTGETPERFTYIAHQDTGRYPEGD	566
QY	557	-----NVMITDEEIKATNPVATERFGTVAVNFQSSSTDPRATGDVHAMGALPGMWQD	609
Db	567	WIONINENLPVTNDNVLPTDPIG---GKTGINY--TNIFNTYGLPLTALNNVP-----	614
QY	610	RDVYLOGPIWAKIPIHTDGHFHPBPLMGAGFLK-----NPPQILIKNTPVPA	656
Db	615	-PVPYNGQIWDKEFDTD-----LKPRLHVNAFPVCCQNNCPQOLEFVKVAPNLT	660
QY	657	N---PPAEFSATKFASFITQYSTQGVSEIEMELQKENSKRWNPEVQYTSNYAKSANDV	713
Db	661	NEYDPDASANMSR-----IYVYSDFWWKGLVKAKLRASHWTWNPICQMSIN-----	707
QY	714	TVDNNGLYTEPRPIG 728	
Db	708	-VDNQFNIV-PSNIG 720	

```

RESULT 15
COAT_PAVHH
ID COAT_PAVHH STANDARD; PRT; 722 AA.
AC P03136;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Coat protein VP1 (Contains: Coat protein VP2).
OS Hamster parvovirus H1.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
OX NCBI_TaxID=10799;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83112183; PubMed=6823009;
RA Rhode S.L. III, Paradiso P.R.;
RT "Parvovirus genome: nucleotide sequence of H-1 and mapping of its
RT genes by hybrid-arrested translation.";
RL J. Virol. 45:173-184 (1983).
CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.

```

```
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

CC -----  
DR EMBL; X01457; CAB57285.1; ALT\_SEQ.  
DR PIR; A03699; VCVV2.  
DR HSSP; P07302; 1MW.  
DR InterPro; IPR001403; Parvo\_coat.  
DR Pfam; PF00740; Parvo\_coat; 1.  
KW Coat protein; Glycoprotein.  
FT CHAIN 1 722 COAT PROTEIN VP1





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 2003, 18:26:09 ; Search time 54.7373 Seconds  
(without alignments)  
2770.518 Million cell updates/sec

Title: US-09-807-802A-13

Perfect score: 3989

Sequence: 1 MAADGYLPDWLEDNLSEGR.....NNGLYTEPRPIGTRYLTRPL 736

Scoring table: BL0SUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_rvirus:\*
- 16: sp\_bacteriaph:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3989	100.0	736	12	Q9WBP8	Q9wbp8 adeno-associ
2	3963	99.3	736	12	O56137	O56137 adeno-associ
3	3511	88.0	736	12	O56139	O56139 adeno-associ
4	3494	87.6	736	12	O65311	O65311 adeno-associ
5	3402.5	85.3	735	12	O56652	O56652 adeno-associ
6	2759.5	69.2	598	12	O56653	O56653 adeno-associ
7	2486.5	62.3	734	12	O41855	O41855 adeno-associ
8	2481.5	62.2	533	12	O92917	O92917 adeno-associ
9	2220	55.7	724	12	O9Y1J1	O9y1j1 adeno-associ
10	2190.5	54.9	732	12	O83290	O83290 muscovy duc
11	2173.5	54.5	732	12	O67666	O67666 goose parvo
12	2168.5	54.4	732	12	O83289	O83289 muscovy duc
13	2165.5	54.3	732	12	O8V395	O8v395 goose parvo
14	2162.5	54.2	732	12	O65444	O65444 barbarie du
15	1990.5	49.9	676	12	O67672	O67672 goose parvo
16	1712	42.9	587	12	O67667	O67667 goose parvo

17	1698	42.6	534	12	O67668	O67668 goose parvo
18	1695	42.5	587	12	O65445	O65445 barbarie du
19	1676	42.0	534	12	O65446	O65446 barbarie du
20	859.5	21.5	703	12	O8QOV5	O8qov5 minute viru
21	715	17.9	179	12	O9WA24	O9wa24 goose parvo
22	715	17.9	179	12	O9W8U7	O9w8u7 duck parvov
23	711	17.8	179	12	O9WN18	O9wn18 duck parvov
24	709	17.8	179	12	O9WN19	O9wn19 duck parvov
25	707	17.7	179	12	O9WN20	O9wn20 goose parvo
26	618	15.5	571	12	O8QOV4	O8qov4 minute viru
27	617	15.5	785	12	O9J0X4	O9j0x4 pig-tailed
28	614.5	15.4	781	12	O9PZT0	O9pzt0 hamster par
29	611	15.3	781	12	O8QVL1	O8qvl1 hamster par
30	608.5	15.3	781	12	O9JGP8	O9jgp8 human parvo
31	606.5	15.2	781	12	O912B8	O912b8 human eryth
32	606.5	15.2	781	12	P89317	P89317 human parvo
33	606.5	15.2	781	12	P89318	P89318 human parvo
34	606.5	15.2	781	12	P89319	P89319 human parvo
35	604.5	15.2	773	12	O913X1	O913x1 human parvo
36	604.5	15.2	781	12	P89316	P89316 human parvo
37	603.5	15.1	781	12	P90221	P90221 human parvo
38	603.5	15.1	781	12	P90222	P90222 human parvo
39	603.5	15.1	781	12	P90223	P90223 human parvo
40	603.5	15.1	781	12	P90224	P90224 human parvo
41	603.5	15.1	781	12	O85117	O85117 human parvo
42	603.5	15.1	781	12	O90200	O90200 human parvo
43	603	15.1	829	12	P87584	P87584 chipmunk pa
44	602.5	15.1	781	12	O9JGS0	O9jgs0 human parvo
45	601.5	15.1	781	12	P89321	P89321 human parvo

ALIGNMENTS

RESULT 1

Q9WBP8 PRELIMINARY; PRT; 736 AA.  
ID Q9WBP8  
AC Q9WBP8;  
DT 01-NOV-1999 (TREMBlrel. 12, Created)  
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE Capsid protein.  
OS adeno-associated virus 1.  
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.  
OX NCBI\_TaxID=85106;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=9214338; PubMed=10196295;  
RA Xiao W., Chirmule N., Berta S.C., McCullough B., Gao G., Wilson J.M.;  
RT "Gene therapy vectors based on adeno-associated virus type 1.";  
RL J. Virol. 73:3994-4003(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Xiao W., Wilson J.M.;  
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF063497; AAD27757.1; -.  
DR InterPro; IPR001403; Parvo\_coat.  
DR Pfam; PF00740; Parvo\_coat; 1.  
SQ SEQUENCE 736 AA; 81375 MW; CFABFB9BD5CD0595 CRC64;

Query Match 100.0%; Score 3989; DB 12; Length 736;  
Best Local Similarity 100.0%; Pred. No. 7.9e-258;  
Matches 736; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MAADGYLPDWLEDNLSEGRWMDLKPAPKPKANQKODDGRGLVLPGYKYLGPFGULD	60
DB	1	MAADGYLPDWLEDNLSEGRWMDLKPAPKPKANQKODDGRGLVLPGYKYLGPFGULD	60
QY	61	KGEFVNADAALAHDKAYDQQLKAGDNPYLRYNHADAERLQEDTSFGNIGRAVFQ	120
DB	61	KGEFVNADAALAHDKAYDQQLKAGDNPYLRYNHADAERLQEDTSFGNIGRAVFQ	120
QY	121	AKKRVLEPLGLVEGAKTAPGKKRPVEQSQEPDSSSGIGKTGQPAKKRLNFGQTGDSE	180

```
|||||
Db 121 AKKRVLEPLGLVEEAKTAPGKKRPVEQSPQEPDSSSGIGTKGQPAKKRLNFGQTGDSE 180
Qy 181 SVDPDQPLGEPPATPAVGPPTMASGGGAPMADNNEGADVGNASGNWCHDSTWLGDRVI 240
Db 181 SVDPDQPLGEPPATPAVGPPTMASGGGAPMADNNEGADVGNASGNWCHDSTWLGDRVI 240
Qy 241 TTSTRTVALPTYNHLYKQISSASTGASNDNHFGYSTPMGYFDENRHHCHFSPRDWQRL 300
Db 241 TTSTRTVALPTYNHLYKQISSASTGASNDNHFGYSTPMGYFDENRHHCHFSPRDWQRL 300
Qy 301 INNMGFRPKRLNFKLFNIQVKEVTNDGVTTIANNLTSTVQVFSDEYQLPYVLGSAHQ 360
Db 301 INNMGFRPKRLNFKLFNIQVKEVTNDGVTTIANNLTSTVQVFSDEYQLPYVLGSAHQ 360
Qy 361 GCLPPFPADVFMIPQYGYLTLNNGSOAVGRSSFYCLEYFSPQMLRTGNFTFSYTFEEVP 420
Db 361 GCLPPFPADVFMIPQYGYLTLNNGSOAVGRSSFYCLEYFSPQMLRTGNFTFSYTFEEVP 420
Qy 421 FHSSYAHQSQSLDRMLNPLIDQYLYLNRTQNGSGAQONKDLFSRGSAPAGMSVQPKMWLP 480
Db 421 FHSSYAHQSQSLDRMLNPLIDQYLYLNRTQNGSGAQONKDLFSRGSAPAGMSVQPKMWLP 480
Qy 481 GPCYRQQRVSKTKTDNNNSNFTWTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMSCV 540
Db 481 GPCYRQQRVSKTKTDNNNSNFTWTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMSCV 540
Qy 541 MIFGESAGASNTALDNVMTDEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMG 600
Db 541 MIFGESAGASNTALDNVMTDEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMG 600
Qy 601 ALPGMWQDRDVYLOGPWAKIPHTDGHFHSPLMGFGLKHPPOILLIKNTPVPANPPA 660
Db 601 ALPGMWQDRDVYLOGPWAKIPHTDGHFHSPLMGFGLKHPPOILLIKNTPVPANPPA 660
Qy 661 EFSATKPFASFITOYSTGQVSVEIEMELQKENSKRWNPEVOYTSNYAKSANVDFTVDNNGL 720
Db 661 EFSATKPFASFITOYSTGQVSVEIEMELQKENSKRWNPEVOYTSNYAKSANVDFTVDNNGL 720
Qy 721 YTEPRPIGTRYLTRPL 736
Db 721 YTEPRPIGTRYLTRPL 736
```

## RESULT 2

```
056137 PRELIMINARY; PRT; 736 AA.
ID 056137;
AC 056137;
DT 01-JUN-1998 (Tremblrel. 06, Created)
DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE Capsid protein VP1.
OS adeno-associated virus 6.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
OX NCBI_TaxID=68558;
RN [1]
RP SEQUENCE FROM N.A.
RA Rutledge E.A., Russell D.W.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF028704; AAB95450.1;
DR InterPro; IPR001403; Parvo_coat.
DR Pfam; PF00740; Parvo_coat_1.
SQ SEQUENCE 736 AA; 81411 MW; 311217A089C565F5 CRC64;
```

Query Match 99.3%; Score 3963; DB 12; Length 736;  
Best Local Similarity 99.2%; Pred. No. 4.3e-256;

```
Matches 730; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
Qy 1 MAADGYLPDWLEDNLSEGIREWDLKPGAPKPKANQKQDDGRGLVPGYKYLGFENGLD 60
Db 1 MAADGYLPDWLEDNLSEGIREWDLKPGAPKPKANQKQDDGRGLVPGYKYLGFENGLD 60
Qy 61 KGEPVNAADAALAEHDKAYDQOLKAGDNPYLRYNHADAEPQERLQEDTSFGNLGRAVFQ 120
Db 61 KGEPVNAADAALAEHDKAYDQOLKAGDNPYLRYNHADAEPQERLQEDTSFGNLGRAVFQ 120
Qy 121 AKKRVLEPLGLVEEAKTAPGKKRPVEQSPQEPDSSSGIGTKGQPAKKRLNFGQTGDSE 180
Db 121 AKKRVLEPLGLVEEAKTAPGKKRPVEQSPQEPDSSSGIGTKGQPAKKRLNFGQTGDSE 180
Qy 181 SVDPDQPLGEPPATPAVGPPTMASGGGAPMADNNEGADVGNASGNWCHDSTWLGDRVI 240
Db 181 SVDPDQPLGEPPATPAVGPPTMASGGGAPMADNNEGADVGNASGNWCHDSTWLGDRVI 240
Qy 241 TTSTRTVALPTYNHLYKQISSASTGASNDNHFGYSTPMGYFDENRHHCHFSPRDWQRL 300
Db 241 TTSTRTVALPTYNHLYKQISSASTGASNDNHFGYSTPMGYFDENRHHCHFSPRDWQRL 300
Qy 301 INNMGFRPKRLNFKLFNIQVKEVTNDGVTTIANNLTSTVQVFSDEYQLPYVLGSAHQ 360
Db 301 INNMGFRPKRLNFKLFNIQVKEVTNDGVTTIANNLTSTVQVFSDEYQLPYVLGSAHQ 360
Qy 361 GCLPPFPADVFMIPQYGYLTLNNGSOAVGRSSFYCLEYFSPQMLRTGNFTFSYTFEEVP 420
Db 361 GCLPPFPADVFMIPQYGYLTLNNGSOAVGRSSFYCLEYFSPQMLRTGNFTFSYTFEEVP 420
Qy 421 FHSSYAHQSQSLDRMLNPLIDQYLYLNRTQNGSGAQONKDLFSRGSAPAGMSVQPKMWLP 480
Db 421 FHSSYAHQSQSLDRMLNPLIDQYLYLNRTQNGSGAQONKDLFSRGSAPAGMSVQPKMWLP 480
Qy 481 GPCYRQQRVSKTKTDNNNSNFTWTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMSCV 540
Db 481 GPCYRQQRVSKTKTDNNNSNFTWTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMSCV 540
Qy 541 MIFGESAGASNTALDNVMTDEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMG 600
Db 541 MIFGESAGASNTALDNVMTDEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMG 600
Qy 601 ALPGMWQDRDVYLOGPWAKIPHTDGHFHSPLMGFGLKHPPOILLIKNTPVPANPPA 660
Db 601 ALPGMWQDRDVYLOGPWAKIPHTDGHFHSPLMGFGLKHPPOILLIKNTPVPANPPA 660
Qy 661 EFSATKPFASFITOYSTGQVSVEIEMELQKENSKRWNPEVOYTSNYAKSANVDFTVDNNGL 720
Db 661 EFSATKPFASFITOYSTGQVSVEIEMELQKENSKRWNPEVOYTSNYAKSANVDFTVDNNGL 720
Qy 721 YTEPRPIGTRYLTRPL 736
Db 721 YTEPRPIGTRYLTRPL 736
```

## RESULT 3

```
056139 PRELIMINARY; PRT; 736 AA.
ID 056139;
AC 056139;
DT 01-JUN-1998 (Tremblrel. 06, Created)
DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE Capsid protein VP1.
OS adeno-associated virus 3B.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
OX NCBI_TaxID=68742;
RN [1]
RP SEQUENCE FROM N.A.
RA Rutledge E.A., Russell D.W.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF028704; AAB95450.1;
DR InterPro; IPR001403; Parvo_coat.
DR Pfam; PF00740; Parvo_coat_1.
SQ SEQUENCE 736 AA; 81411 MW; 311217A089C565F5 CRC64;
```





QY 720 LYTEPRPIGTRYLTRPL 736  
Db 720 YSEPRPIGTRYLTRNL 736

RESULT 5

056652 PRELIMINARY; PRT; 735 AA.  
ID 056652  
AC 056652;  
DT 01-JUN-1998 (TREMBlrel. 06, Created)  
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE Major coat protein VP1.  
OS Adeno-associated virus 2 (AAV2).  
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.  
OX NCBI\_TaxID=10804;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95088582; Pubmed=7996133;  
RA Ruffing M., Heid H., Kleinschmidt J.A.;  
RT "Mutations in the carboxy terminus of adeno-associated virus 2 capsid proteins affect viral infectivity: lack of an RGD integrin-binding motif."  
RT J. Gen. Virol. 75:0-0(0).  
RL [2]  
RN SEQUENCE FROM N.A.  
RP Berns K.I., Bohenzky R.A., Cassinotti P., Colvin D., Donahue B.A.,  
RA Dull T., Horer M., Kleinschmidt J.A., Ruffing M., Snyder R.O.,  
RA Tratschin J.-D., Weitz M.;  
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF043303; AAC03780.1; -  
DR InterPro; IPR001403; Parvo\_coat.  
DR Pfam; PF00740; Parvo\_coat; 1.  
KW Coat protein.  
FT VARIANT 76 76 D -> V.  
FT VARIANT 553 553 D -> N.  
FT VARIANT 567 567 T -> N.  
FT VARIANT 677 678 QV -> HV.  
FT VARIANT 710 710 V -> R.  
SQ SEQUENCE 735 AA; 81944 MW; 980BEEF46908390B CRC64;

Query Match 85.3%; Score 3402.5; DB 12; Length 735;  
Best Local Similarity 83.3%; Pred. No. 1.1e-218;  
Matches 613; Conservative 51; Mismatches 71; Indels 1; Gaps 1;

QY 1 MAADGYLPDWLEDNLSGIREWMDLKGAPKPKANQOKODD3RGLVLPGYKYLGPFGD 60  
Db 1 MAADGYLPDWLEDTLSEGIROWWKLKPGPPPKPAERHKDD3RGLVLPFGYKYLGPFGD 60  
QY 61 KGEPVNADAAALHDKAYDQOLKAGDNPRYLRYNHADAERLQEDTSFGCNLGRAVFO 120  
Db 61 KGEPVNEADAAALHDKAYDRQLDSGDNPRYLKYNHADAERLQEDTSFGCNLGRAVFO 120  
QY 121 AKKRVLEPLGLVEEAKTAPGKKRPVEQSPOEPDSSSGIGTGOQPAKKRLNFGQTGDS 180  
Db 121 AKKRVLEPLGLVEEPVKTAPEGKRPVEHSPVEPDSSSGTGKAGQOPAKRRLNFGQTGDAD 180  
QY 181 SVDPDQPLGPPAPSGLTNTMATGSGAPMADNNEGADGVNAGNWHCDSTWMDRVI 240  
Db 181 SVDPDQPLGPPAPSGLTNTMATGSGAPMADNNEGADGVNAGNWHCDSTWMDRVI 240  
QY 241 TTSTRWALPTYNHLYKQISSASTGASNDNHYFGYSTPMGYFDENRFCHFSPRDQRL 300  
Db 241 TTSTRWALPTYNHLYKQISSQS-GASNDNHYFGYSTPMGYFDENRFCHFSPRDQRL 299  
QY 301 INNNWGFRPKRLNFKLNIQVKEVTNDGVTITANNLTSTVQVFSDSRYQLPYVLSAHQ 360  
Db 301 INNNWGFRPKRLNFKLNIQVKEVTQNDGVTITANNLTSTVQVFTDSEYQLPYVLSAHQ 359  
QY 361 GCLPPFPADVEMIPQYGYLLTLNNGSQAVGRSSFYCLEYFSPQMLRTGNFTFSYTFEEVP 420  
Db 360 GCLPPFPADVEMIPQYGYLLTLNNGSQAVGRSSFYCLEYFSPQMLRTGNFTFSYTFEDVP 419

QY 421 FHSSYAHQSOLDRLMPLIDQYLYLNRTONOSGSAQNKDLFSRGSFAGMSVQPKNMLP 480  
Db 420 FHSSYAHQSOLDRLMPLIDQYLYLSRTNTPSGTTQSRLOFSQAGASDIRDQSRNMLP 479  
QY 481 GPCYRQQRVSKTKTDNNNSNFTWTGASKYNLNGRESTINPGTAMASHKODEDKFFPMGCV 540  
Db 480 GPCYRQQRVSKTSADNNNSNFTWTGATKYHLNGRDSLVPNPGPAMASHKODEEKFPPQSGV 539  
QY 541 MIFGKESAGASNTALDNVMTDEEIKATNPVATERFGTAVAVNFQSSSTDPATGDVHAMG 600  
Db 540 LIFGKQSEKTNVDIEKVMITDEEIRITNPVATEQYGSVSTNLQRGNRQAATADVNTQG 599  
QY 601 ALPGMWQDRDVYLQGPWAKIPHTDGHFHPSPLMGFGGLKNPPEQILIKNTPVPANPPA 660  
Db 600 VLPGMWQDRDVYLQGPWAKIPHTDGHFHPSPLMGFGGLKNPPEQILIKNTPVPANPST 659  
QY 661 EFSATKFAFITQYSTGVSEIEMELQKENSKRWNPEVOYTSNYAKSANVDFTVDNNGL 720  
Db 660 TFSAKTFAFITQYSTGVSEIEMELQKENSKRWNPEIQTSNYKSVNVDFTVDITNGV 719  
QY 721 YTEPRPIGTRYLTRPL 736  
Db 720 YSEPRPIGTRYLTRNL 735

RESULT 6

056653 PRELIMINARY; PRT; 598 AA.  
ID 056653  
AC 056653;  
DT 01-JUN-1998 (TREMBlrel. 06, Created)  
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE Major coat protein VP2.  
OS Adeno-associated virus 2 (AAV2).  
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.  
OX NCBI\_TaxID=10804;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95088582; Pubmed=7996133;  
RA Ruffing M., Heid H., Kleinschmidt J.A.;  
RT "Mutations in the carboxy terminus of adeno-associated virus 2 capsid proteins affect viral infectivity: lack of an RGD integrin-binding motif."  
RT J. Gen. Virol. 75:0-0(0).  
RL [2]  
RN SEQUENCE FROM N.A.  
RP Berns K.I., Bohenzky R.A., Cassinotti P., Colvin D., Donahue B.A.,  
RA Dull T., Horer M., Kleinschmidt J.A., Ruffing M., Snyder R.O.,  
RA Tratschin J.-D., Weitz M.;  
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF043303; AAC03778.1; -  
DR InterPro; IPR001403; Parvo\_coat.  
DR Pfam; PF00740; Parvo\_coat; 1.  
KW Coat protein.  
FT VARIANT 416 416 D -> N.  
FT VARIANT 430 430 T -> N.  
FT VARIANT 540 541 QV -> HV.  
FT VARIANT 573 573 V -> R.  
SQ SEQUENCE 598 AA; 66619 MW; 070811ED9368E934 CRC64;

Query Match 69.2%; Score 2759.5; DB 12; Length 598;  
Best Local Similarity 82.8%; Pred. No. 6.4e-176;  
Matches 495; Conservative 43; Mismatches 59; Indels 1; Gaps 1;

QY 139 APGKKRPVEQSPOEPDSSSGIGTGOQPAKKRLNFGQTGDSVDPQPLGEPATPAAV 198  
Db 2 APGKKRPVEHSPVEPDSSSGTGKAGQOPAKRRLNFGQTGDADSVDPQPLGPPAPPSGL 61  
QY 199 GPTTMAAGGAPMADNNEGADGVNAGNWHCDSTWLGDRVITSTRWALPTYNHLYK 258  
Db 62 GTNTMATGSGAPMADNNEGADGVNAGNWHCDSTWMDRVIITSTRWALPTYNHLYK 121

QY 259 QISSASTGASNDNHYPGYSTPMGYFEDFNRFHCHFSPRDMORLLINNMGFRPKRLNFKLFN 318  
 DB 122 QISSQS-GASNDNHYPGYSTPMGYFEDFNRFHCHFSPRDMORLLINNMGFRPKRLNFKLFN 180  
 QY 319 IQVKEVTNDGVTTIANNLTSTVOVFSDEYQLPYVLGSAHQGLPFPADVEMIPOYGY 378  
 DB 181 IQVKEVTNDGVTTIANNLTSTVOVFSDEYQLPYVLGSAHQGLPFPADVEMIPOYGY 240  
 QY 379 LTLNNGSQAVGRSSFYCYLEYFSPQMLRTGNFTFSYTFEEVPHSSYAHQSOLDRLMNP 438  
 DB 241 LTLNNGSQAVGRSSFYCYLEYFSPQMLRTGNFTFSYTFEEVPHSSYAHQSOLDRLMNP 300  
 QY 439 IDOYLTYLNRTQNSGSAQNKDLFSRGSFAGMSVQPKNWLPGCYRQORVSKTKTDNN 498  
 DB 301 IDOYLTYLNRTQNSGSAQNKDLFSRGSFAGMSVQPKNWLPGCYRQORVSKTKTDNN 360  
 QY 499 SNFTWTGASKNLNGRESIINPGTAMASHKDEDEKFPFMSGVMIFGKESAGASNTALDNV 558  
 DB 361 SEYSWTGATKHLNGRDSLVPNGPAMASHKDEDEKFPFMSGVMIFGKESAGASNTALDNV 420  
 QY 559 MITDEEIKATNPVATERFGTAVAVNFSSSTDPAATGVDHAMGALPGMWQDRDVLQGP 618  
 DB 421 MITDEEIKATNPVATERFGTAVAVNFSSSTDPAATGVDHAMGALPGMWQDRDVLQGP 480  
 QY 619 WAKIPHTDGHFHPSPILMGFGLKPNPPQILIKNTVPANPAEPFSAATKFAFITQYSTGQ 678  
 DB 481 WAKIPHTDGHFHPSPILMGFGLKPNPPQILIKNTVPANPAEPFSAATKFAFITQYSTGQ 540  
 QY 679 VSVEIEMELQENSKRWNPVOYTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 736  
 DB 541 VSVEIEMELQENSKRWNPVOYTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 598

RESULT 7

QY 041855 PRELIMINARY; PRT; 734 AA.  
 AC 041855;  
 DT 01-JAN-1998 (TREMBlrel. 05, Created)  
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE Capsid.  
 OS adeno-associated virus 4.  
 OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.  
 OX NCBI\_TaxID=57579;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC VR-646;  
 RX MEDLINE=97404695; PubMed=9261407;  
 RA Chiorini J.A., Yang L., Liu Y., Safer B., Kotin R.M.;  
 RT "Cloning of adeno-associated virus type 4 (AAV4) and generation of  
 recombinant AAV4 particles."  
 RL J. Virol. 71:6823-6833(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC VR-646;  
 RA Chiorini J.A., Yang L., Kotin R.M., Safer B.;  
 RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U89790; AAC58045.1; -  
 DR InterPro; IPR001403; Parvo\_coat.  
 DR Pfam; PF00740; Parvo\_coat; 1.  
 SQ SEQUENCE 734 AA; 80639 MW; 616CC27A777BBE6F CRC64;

Query Match 62.3%; Score 2486.5; DB 12; Length 734;  
 Best Local Similarity 63.5%; Pred. No. 1.4e-157;  
 Matches 475; Conservative 81; Mismatches 161; Indels 31; Gaps 10;

QY 4 DGYLPWMLDNLSEGIRESWMDLPGAPKPKANQOKQDDGRGLVLPGYKYLGPFGNLDKGE 63  
 DB 3 DGYLPWMLDNLSEGIRESWMDLPGAPKPKANQOKQDDGRGLVLPGYKYLGPFGNLDKGE 62  
 QY 64 PVNAADAAALEHDKAYDQOLKAGDNPYLRYNHADAEPFOERLQEDTSGFNLGRAVFOAKK 123  
 DB 63 PVNAADAAALEHDKAYDQOLKAGDNPYLRYNHADAEPFOERLQEDTSGFNLGRAVFOAKK 122

QY 124 RVLEPLGLVEGAKTAPGKKRPEVQSPQEPDSSSGIGKTGOOPAKKRLNF-GOTGDSSEV 182  
 DB 123 RVLEPLGLVEGAKTAPGKKRPEVQSPQEPDSSSGIGKTGOOPAKKRLNF-GOTGDSSEV 182  
 QY 183 PDPQPLGEPPATPAVGPFTTASGGGAPMADNNEGADGVNAGNWHCDSTWLGDRVIT 242  
 DB 183 PDPQPLGEPPATPAVGPFTTASGGGAPMADNNEGADGVNAGNWHCDSTWLGDRVIT 236  
 QY 243 STRTVALPTNNHLYKQISSASTGASNDNHYPGYSTPMGYFEDFNRFHCHFSPRDMORLIN 302  
 DB 237 STRTVALPTNNHLYKQISSASTGASNDNHYPGYSTPMGYFEDFNRFHCHFSPRDMORLIN 292  
 QY 303 NNMGFRPKRLNFKLFNIQVKEVTNDGVTTIANNLTSTVOVFSDEYQLPYVLGSAHQGC 362  
 DB 293 NNMGFRPKRLNFKLFNIQVKEVTNDGVTTIANNLTSTVOVFSDEYQLPYVLGSAHQGC 352  
 QY 363 LPPFPADVEMIPOYGY--LTLNNGSQAVGRSSFYCYLEYFSPQMLRTGNFTFSYTFEEV 419  
 DB 353 LPPFPADVEMIPOYGY--LTLNNGSQAVGRSSFYCYLEYFSPQMLRTGNFTFSYTFEEV 412  
 QY 420 PFHSSYAHQSOLDRLMNPIDQYLTYLNRTQNSGSAQNKDLFSRGSFAGMSVQ 475  
 DB 413 PFHSSYAHQSOLDRLMNPIDQYLTYLNRTQNSGSAQNKDLFSRGSFAGMSVQ 469  
 QY 476 KMWLPGCYRQORVSKTKTDNNNSNFTWTGAS--KY--NNGRESIINPGTAMASHK 528  
 DB 470 KMWLPGCYRQORVSKTKTKTDNNNSNFTWTGAS--KY--NNGRESIINPGTAMASHK 527  
 QY 529 DDEDEKFPFMSGVMIFGKESAGASNTALDNVMITDEEIKATNPVATERFGTAVAVNFSS 588  
 DB 528 DDEDEKFPFMSGVMIFGKESAGASNTALDNVMITDEEIKATNPVATERFGTAVAVNFSS 586  
 QY 589 TDPATGVDHAMGALPGMWQDRDVLQGPWAKIPHTDGHFHPSPILMGFGLKPNPPQIL 648  
 DB 587 TDPATGVDHAMGALPGMWQDRDVLQGPWAKIPHTDGHFHPSPILMGFGLKPNPPQIL 646  
 QY 649 IKNTVPANPAEPFSAATKFAFITQYSTGQVSVEIEMELQENSKRWNPVOYTSNYAKS 708  
 DB 647 IKNTVPANPAEPFSAATKFAFITQYSTGQVSVEIEMELQENSKRWNPVOYTSNYAKS 706  
 QY 709 ANVDFTVDNNGLYTEPRPIGTRYLTRPL 736  
 DB 707 ANVDFTVDNNGLYTEPRPIGTRYLTRPL 734

RESULT 8

QY 092917 PRELIMINARY; PRT; 533 AA.  
 AC 092917;  
 DT 01-NOV-1998 (TREMBlrel. 08, Created)  
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE Major coat protein VP3.  
 OS Adeno-associated virus 2 (AAV2).  
 OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.  
 OX NCBI\_TaxID=10804;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=95088582; PubMed=7996133;  
 RX Ruffing M., Heid H., Kleinschmidt J.A.;  
 RA "Mutations in the carboxy terminus of adeno-associated virus 2 capsid  
 proteins affect viral infectivity: lack of an RGD integrin-binding  
 motif."  
 RL J. Gen. Virol. 75:0-0(0).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC Berns K.I., Bohenzky R.A., Cassinotti P., Colvin D., Donahue B.A.,  
 RA Dull T., Horer M., Kleinschmidt J.A., Ruffing M., Snyder R.O.,  
 RA Tratschin J.-D., Weitz M.;  
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF043303; AAC03779.1; -  
 DR InterPro; IPR001403; Parvo\_coat.





OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.  
OX NCBI\_TaxID=37325;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=89384/France;  
RX MEDLINE=96406928; PubMed=8811015;  
RA Le Gall-Recul G., Jestin V., Chagnaud P., Blanchard P., Jestin A.;  
RT "Expression of muscovy duck parvovirus capsid proteins (VP2 and VP3)  
in a baculovirus expression system and demonstration of immunity  
induced by the recombinant protein.";  
RL J. Gen. Virol. 77:2159-2163(1996).  
DR EMBL; 268272; CA92575.1; -  
DR InterPro; IPR001403; Parvo\_coat.  
DR Pfam; PF00740; Parvo\_coat; 1.  
FT CHAIN 146 732 VP2 CAPSID PROTEIN.  
FT CHAIN 199 732 VP3 CAPSID PROTEIN.  
SQ SEQUENCE 732 AA; 81364 MW; DE70DCCAB215F4E2 CRC64;

Query Match 54.9%; Score 2190.5; DB 12; Length 732;  
Best Local Similarity 55.7%; Pred. No. 7.9e-138;  
Matches 422; Conservative 88; Mismatches 184; Indels 63; Gaps 13;

QY 9 DWLEDNLSGIREWMDLKPAPKPKANQOKQ-----DDGRGLVLPGYKYLGPFN 57  
DB 10 DWVE-----TAAASWRHLKAGAPKPKSNQSQSVSTDRKPKRKNNRGFVLPGYKYLGPFN 65  
QY 58 GLDKGEPVNAADAALAEHDKAYDQOLKAGDNPYLRYNHADAERLQEDTSFGNLGRA 117  
DB 66 GLDKGEPVNAKADSVALEHDKAYDQOLKAGDNPYIKFNHADQEFIDNLQDTSFGNLGRA 125  
QY 118 VFQAKKRVLEPLGLVEEGAKTAPGKKRPVEQSPQEPDSSSGIGTKGQOPAKKRLNFGQTG 177  
DB 126 VFQAKKRILEPLGLVEEPVNTAPAKK-----SSG-KLTDHYPIVKPKLSE-- 170  
QY 178 DSESVPDPQLG-----EPPATPAVGPPTMASGGGAPMADNNEGADVGNASG 226  
DB 171 --ENSPPSPNSGGEASAATEGSEVAP-----NMAEGSGAMGDSAGADGVGNASG 222  
QY 227 NMHCDSTWLGDRIITSTRTWALPTVNNHLKYQISSASTGASNDNHFGYSTPWFYDFPN 286  
DB 223 NMHCDSQWLGDVITTKTRTWLPSYNNHIYKAITSGTNPSDN--TQYAGYSTPWFYDFPN 281  
QY 287 RFHGHSPRDWQRLINNNWGFPRKRLNFKLNIQVKEVTNDGVTTIANNLTSTVOVPSD 346  
DB 282 RFHGHSPRDWQRLINNNWGFPRKRLNFKLNIQVKEVTNDGVTTIANNLTSTVOVPSD 341  
QY 347 SEYQLPYVLGSAHQGLPPFPADVEMIPQYGYLTN--NGSQAVGRSSFYCLEYFPGM 403  
DB 342 NEHQLPYVLGSAHQGLPPFPADVEMIPQYGYLTN--NGSQAVGRSSFYCLEYFPGM 401  
QY 404 LRTGNNTFTSYTFEEVPFHSSYAHQSGLDRLMPLIDQYLYLNRTQNGSGSAQNKDLF 463  
DB 402 LRTGNNTFTSYTFEEVPFHSSYAHQSGLDRLMPLIDQYLYLNRTQNGSGSAQNKDLF 455  
QY 464 SRGSPAGMSVQPKNWLPGPCYRQORVSKTK--TDNNNSNFTWTGASKYNLNGRESINPG 521  
DB 456 KKAIVKAGAFGAMGRNWLPGPKLDPQVRAYSGGTDNYANWSIWSKGNKVFLLKDRYLLQPG 515  
QY 522 TAMASHKDEDEKFFPMGVMIFGKE--SAGASNTALDNVMTDEEIKATNPVATERFGT 579  
DB 516 PVAATHTEHQASSVPAQNIIGIAKDPYRSSTLAGISDIMVDEQEIAPTNVGWRPYGL 575  
QY 580 VAVNFQSSSTDPATGVDVHMGALPGWVQODRDVYLQGPIMAKIPHTDGHFSPMLMGFG 639  
DB 576 TVTNEQNTTAPTAELLEVGLALPGWVQNRDIYLGPIWAKIPKTDGKFPSPNLGGFG 635  
QY 640 LKNPPQILIKNTVPANPPAEFSATKPFASFITQYSTGQVSEIEMELQKENSKRANPEV 699  
DB 636 LKNPPQVFIKNTVPADPPLLEYVNOQKNSYITQYSTGQCTVEMWELRKENSKRWNPFI 695  
QY 700 QYTSNTAKSANVDFTVNNGLYTEPPRIGTRYLTRPL 736  
DB 696 QYTSNFGNRTSTMFAPIVNETGGYVEDRLIGTRYLTQNL 732

RESULT 11  
ID Q67666 PRELIMINARY; PRT; 732 AA.  
AC Q67666;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE VP1.  
GN VP1.  
OS goose parvovirus.  
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.  
OX NCBI\_TaxID=38251;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VIRULENT B;  
RX MEDLINE=96010229; PubMed=7571426;  
RA Zadori Z., Stefancsik R., Rauch T., Kisary J.;  
RT "Analysis of the complete nucleotide sequences of goose and muscovy  
RT duck parvoviruses indicates common ancestral origin with adeno-  
RT associated virus 2.";  
RL Virology 212:562-573(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VIRULENT B;  
RA Zadori Z.;  
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U25749; AAA83230.1; -  
DR InterPro; IPR001403; Parvo\_coat.  
DR Pfam; PF00740; Parvo\_coat; 1.  
SQ SEQUENCE 732 AA; 81340 MW; 339507C61D47B52C CRC64;

Query Match 54.5%; Score 2173.5; DB 12; Length 732;  
Best Local Similarity 54.7%; Pred. No. 1.1e-136;  
Matches 414; Conservative 108; Mismatches 176; Indels 59; Gaps 14;

QY 9 DWLEDNLSGIREWMDLKPAPKPKANQOKQ-----DDGRGLVLPGYKYLGPFN 57  
DB 6 DSFEWYETAASWRHLKAGAPKPKSNQSQSVSPDRPKRDNNGFVLPGYKYLGPFN 65  
QY 58 GLDKGEPVNAADAALAEHDKAYDQOLKAGDNPYLRYNHADAERLQEDTSFGNLGRA 117  
DB 66 GLDKGEPVNAKADSVALEHDKAYDQOLKAGDNPYIKFNHADQEFIDNLQDTSFGNLGRA 125  
QY 118 VFQAKKRVLEPLGLVEEGAKTAPGKKR-----PVEQSPQ-EPDSSSGIGTKGQOPAK 168  
DB 126 VFQAKKRILEPLGLVEDPVTNAPAKNTGKLTLDHYPVVKKPKLTBEVSAGGSSAVQDG- 184  
QY 169 KRLNFGQTDSESVDPDPLGEPATPAVGPPTMASGGGAPMADNNEGADVGNASGNV 228  
DB 185 -----GATAE-----GTEP-----VAASEMAEGGAGMGDSSGADGVGNASGNV 224  
QY 229 HCDSTWLGDRIITSTRTWALPTVNNHLKYQISSASTGASNDNH--YFGYSTPWFYDFPN 286  
DB 225 HCDSQMWGNTVITTKTRTWLPSYNNHIYKAITSGTNPSDN--TQYAGYSTPWFYDFPN 281  
QY 287 RFHGHSPRDWQRLINNNWGFPRKRLNFKLNIQVKEVTNDGVTTIANNLTSTVOVPSD 346  
DB 282 RFHGHSPRDWQRLINNNWGFPRKRLNFKLNIQVKEVTNDGVTTIANNLTSTVOVPSD 341  
QY 347 SEYQLPYVLGSAHQGLPPFPADVEMIPQYGYLTN--NGSQAVGRSSFYCLEYFPGM 403  
DB 342 DEHQLPYVLGSAHQGLPPFPADVEMIPQYGYLTN--NGSQAVGRSSFYCLEYFPGM 401  
QY 404 LRTGNNTFTSYTFEEVPFHSSYAHQSGLDRLMPLIDQYLYLNRTQNGSGSAQNKDLF 463  
DB 402 LRTGNNTFTSYTFEEVPFHSSYAHQSGLDRLMPLIDQYLYLNRTQNGSGSAQNKDLF 455  
QY 464 SRGSPAGMSVQPKNWLPGPCYRQORV-SKTKTDNNNSNFT--WTGASKYNLNGRESINPG 521  
DB 456 KKAIVKAGYGTMRNWLPGPKLDPQVRAYTGTGTDNYANWSIWSKGNKVFLLKDRYLLQPG 515





QY 347 SEXQLPYVLGSAHOGCLPPFPADVFMIPQGYLTLN--NGSOAVGRSSFYCLEYFPSQM 403  
 Db 342 DEHQLPYVLGSAHOGCLPPFPADVFMIPQGYLTLN--NGSOAVGRSSFYCLEYFPSQM 401  
 QY 404 LRTGNFTFSYTFEEVPHSSYAHOSLDRLMPLIDQYLYLNRTONOGSAQNKDLLE 463  
 Db 402 LRTGNFTFSYTFEEVPHSSYAHOSLDRLMPLIDQYLYLNRTONOGSAQNKDLLE 455  
 QY 464 SRGSPAGMSVQPKNMLPGPCYRQORV-SKTKTDNNNSNFT-WTGASKYNLNGRESIINPG 521  
 Db 456 KKAIVKAGAYGTMRNMLPGPKLLDQVRAYTGTDTNANWIMSNKGNKVLKDRQYLLQPG 515  
 QY 522 TAMASHKDEDEKFFPMGVMIFGKE--SAGASNTALDNVMTDEEIKATNPVATERFGT 579  
 Db 516 PVSAHTTKVEASSIPAONILGLADPYRSGSTTAGISDIMVTDEQEVAFPTNGVGMKPYGK 575  
 QY 580 VAVNFQSSSTDPAATGVHAMGALPGMWQDRDYLQGPIMAKIPHTDGHFHPSPMLMGFG 639  
 Db 576 TVTNEQNTTTAPTSSDLVGLALPGMWQNRDIYLGPIWAKIPKTDGKHPSPNLGGFG 635  
 QY 640 LKNPPOILLIKNTVPANPAPAEFSATKFAFITQYSTGQVSEIEMELQKENSKRANPEV 699  
 Db 636 LYNPPPOVFIKNTVPADPPEVYHOKMNSYITQYSSGQCTIVEMWELRKENSKRANPEI 695  
 QY 700 QYTSNYAKSANDFTVDNNGLYTEPRPIGTRYLTRPL 736  
 Db 696 QYTSNFSDRSTIMFAPNETGGYIEDRLIGTRYLTQNL 732

RESULT 14

Q65444 PRELIMINARY; PRT; 732 AA.  
 AC Q65444;  
 DT 01-NOV-1996 (Tremblrel. 01, Created)  
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
 DE Capsid protein VP.  
 GN VP.  
 OS Barbarie duck parvovirus.  
 OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.  
 OX NCBI\_TaxID=39118;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FM;  
 RX MEDLINE=96010229; PubMed=7571426;  
 RA Zadori Z., Stefancsik R., Rauch T., Kisary J.;  
 RT "Analysis of the complete nucleotide sequences of goose and muscovy  
 RT duck parvoviruses indicates common ancestral origin with adeno-  
 RT associated virus 2."  
 RL Virology 212:562-573(1995).  
 DR EMBL; U22967; AAA83225.1;  
 DR InterPro; IPR001403; Parvo\_coat.  
 DR Pfam; PF00740; Parvo\_coat; 1.  
 DR SEQUENCE 732 AA; 81314 MW; D92360596E2D2C05 CRC64;

Query Match 54.2%; Score 2162.5; DB 12; Length 732;  
 Best Local Similarity 55.1%; Pred. No. 5.9e-136;  
 Matches 417; Conservative 90; Mismatches 187; Indels 63; Gaps 13;

QY 9 DWLENDLSEGIKRWMDLPGAPKPKANQOKQ-----DDGRGLVLPGYKYLGPFN 57  
 Db 10 DWYF---TAASWRHLKAGAPKPKANQOKQ-----DDGRGLVLPGYKYLGPFN 65  
 QY 58 GLDKGEFVNAAADAALEHDKAYDQQLKAGDNPYLRYNHADAFAOERLQEDTSFGNLGRA 117  
 Db 66 GLDKGPPVNKADSVALEHDKAYDQQLKAGDNPYIKFKHADQEFIDNLQGDTSFGNLGRA 125  
 QY 118 VFOAKKRVLEPLGLVEEGAKTAPGKRPVEQSPQEPDSSSGIGTKGQOPAKKRLNFGQTG 177  
 Db 126 VFOAKKRVLEPLGLVEEGAKTAPGKRPVEQSPQEPDSSSGIGTKGQOPAKKRLNFGQTG 170  
 QY 178 DSESVDPDQPLG-----EBPATPAVGPPTMASGGGAPMADNNEGADGVGNASG 226

Db 171 --ENSPSPNSNGGEASAATGSEFVAAP-----NMAEGSGAMGDSAGAGDGVGNASG 222  
 QY 227 NWHCDSTWLGDRVITTSRTWALPTYNHLYKQISSASTGASNDNHFGYSTPWGFDEN 286  
 Db 223 NWHCDSTWLGDRVITTSRTWALPTYNHLYKQISSASTGASNDNHFGYSTPWGFDEN 281  
 QY 287 RPHCHFSPRDWORLNNWGFPRKRLNFKLFNIQYKEVTTNDGVTIANLSTVQVFS 346  
 Db 282 RPHCHFSPRDWORLNNWGFPRKRLNFKLFNIQYKEVTTNDGVTIANLSTVQVFS 341  
 QY 347 SEXQLPYVLGSAHOGCLPPFPADVFMIPQGYLTLN--NGSOAVGRSSFYCLEYFPSQM 403  
 Db 342 NEHQLPYVLGSAHOGCLPPFPADVFMIPQGYLTLN--NGSOAVGRSSFYCLEYFPSQM 401  
 QY 404 LRTGNFTFSYTFEEVPHSSYAHOSLDRLMPLIDQYLYLNRTONOGSAQNKDLLE 463  
 Db 402 LRTGNFTFSYTFEEVPHSSYAHOSLDRLMPLIDQYLYLNRTONOGSAQNKDLLE 455  
 QY 464 SRGSPAGMSVQPKNMLPGPCYRQORVSKTK--TDNNNSNFTWTGASKYNLNGRESIINPG 521  
 Db 456 KKAIVKAGAYGTMRNMLPGPKLLDQVRAYTGTDTNANWIMSNKGNKVLKDRQYLLQPG 515  
 QY 522 TAMASHKDEDEKFFPMGVMIFGKE--SAGASNTALDNVMTDEEIKATNPVATERFGT 579  
 Db 516 PVATHTTEDQASSVPAONILGLADPYRSGSTTAGISDIMVTDEQEIAPTNGVWRPYGL 575  
 QY 580 VAVNFQSSSTDPAATGVHAMGALPGMWQDRDYLQGPIMAKIPHTDGHFHPSPMLMGFG 639  
 Db 576 TVTNEQNTTTAPTSSDLVGLALPGMWQNRDIYLGPIWAKIPKTDGKHPSPNLGGFG 635  
 QY 640 LKNPPOILLIKNTVPANPAPAEFSATKFAFITQYSTGQVSEIEMELQKENSKRANPEV 699  
 Db 636 LKNPPOVFIKNTVPADPPEVYHOKMNSYITQYSSGQCTIVEMWELRKENSKRANPEI 695  
 QY 700 QYTSNYAKSANDFTVDNNGLYTEPRPIGTRYLTRPL 736  
 Db 696 QYTSNFSNRTSTIMFAPNETGGYIEDRLIGTRYLTQNL 732

RESULT 15

Q67672 PRELIMINARY; PRT; 676 AA.  
 AC Q67672;  
 DT 01-NOV-1996 (Tremblrel. 01, Created)  
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE VPI (Fragment).  
 OS goose parvovirus.  
 OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.  
 OX NCBI\_TaxID=38251;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SHM 319;  
 RX MEDLINE=95343541; PubMed=7618268;  
 RA Brown K.E., Green S.W., Young N.S.;  
 RT "Goose parvovirus--an autonomous member of the dependovirus genus?";  
 RL Virology 210:283-291(1995).  
 DR EMBL; U34761; AAA75286.1;  
 DR InterPro; IPR001403; Parvo\_coat.  
 DR Pfam; PF00740; Parvo\_coat; 1.  
 DR NON TER 676 676  
 DR SEQUENCE 676 AA; 74800 MW; F28A0B6A115A417F CRC64;

Query Match 49.9%; Score 1990.5; DB 12; Length 676;  
 Best Local Similarity 54.6%; Pred. No. 1.6e-124;  
 Matches 382; Conservative 95; Mismatches 163; Indels 59; Gaps 14;

QY 9 DWLENDLSEGIKRWMDLPGAPKPKANQOKQ-----DDGRGLVLPGYKYLGPFN 57  
 Db 6 DSFEWYETAASWRHLKAGAPKPKANQOKQ-----DDGRGLVLPGYKYLGPFN 65  
 QY 58 GLDKGEFVNAAADAALEHDKAYDQQLKAGDNPYLRYNHADAFAOERLQEDTSFGNLGRA 117

Db 66 GLDKGPPVANKADSVALEHDKAYDLOLKAGDNPIKFNHADQDFIDSLQDDHSGGNLGA 125  
Qy 118 VFOAKKRVLEPLGLVEEGAKTAPGKR-----PVEQSPQ-EPDSSGIGKTGOQPAK 168  
Db 126 VFOAKKRILEPGLVEEPINTAPAKONTGLTDHYPVVKKPKLTEEVSAGGSSAVODG- 184  
Qy 169 KRLNFGQTGDSSEVPDPQPLGEPATPAAVGPTTMASSGGA PMADNNEGADGVNASGNW 228  
Db 185 -----GATAE-----GTEP-----VAASEMAEGGGAMGDSGGADGVNASGNW 224  
Qy 229 HCDSTWLGDRVITTTSTRTWALPTYNNHLYKQISSASTGASNDNH--YFGYSTPMGYFDEN 286  
Db 225 HCDSQWMGNVTITKTRTWLPSYNNHIYKAITSS--GTSQDANQVAGYSTPMGYFDEN 281  
Qy 287 RFHCHFSPRDWQRLINNMGFRPKRLNFKLNIQVKEVTNDZVTIANNLSTVQVFS 346  
Db 282 RFHCHFSPRDWQRLINNHWGIRPKSLKFKIFNVQVKEVTTQDQTKITANNLSTIQVFTD 341  
Qy 347 SEYQLPYVLGSAHQCLPPFPADVFMIPOYGLTLN--NGSQAVGRSSFYCLEYFPSOM 403  
Db 342 DEHQLPYVLGSAETEGTMPFPSPDVYALPQDGYCTWHTNQNGARFNDRSAFYCLEYFPSOM 401  
Qy 404 LRTGNNTFSYTEEVEPFHSSYAHSQSLDRLMNP LIDQYLYLNRTONQSGSAQNKDLF 463  
Db 402 LRTGNNFETTFDEEVEPFHSMFAHSQDLRLMSPLVDQYLMNFNEV-DSSRNAQ-----F 455  
Qy 464 SRGSPAGMSVQPKWLPGPCYRQQRV-SKTKTDNNSNFT-WTGASKYNLNGRESIINPG 521  
Db 456 KKA VKAGYGTGRNWLPGPKFLDQVRVAYTGTGTDNYANWNIMWSNGKNVNLKDRQYLLQPG 515  
Qy 522 TAMASHKDEDEKFFPMSCVMI FGKE--SAGASNTALDNMTDEEIKATNPVATERFGT 579  
Db 516 PVSAHTEGEASSIPAQNILGI AKDPYRSGSTAGISDIMWTDQEVAPTNQVGMKPYGR 575  
Qy 580 VAVNFQSSSTDPATG DVHAMGALPGMWQDRDVLQGP I WAKIPHTDGHFHPSPLMGFG 639  
Db 576 TVTNEQNTTAPITSSDLVDLGLALPGMWQNRDIYLOGP I WAKIPKTDGKHPSPNLRGFG 635  
Qy 640 LKNPPQILIKNTPVANPPAEFSATKFASTTOYSTGO 678  
Db 636 LHNPPQOVFIKNTPVADPFVEYVHQKMSYITTOYSTGO 674

Search completed: July 17, 2003, 18:34:29  
Job time : 58.7373 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 2003, 18:20:14 ; Search time 41.664 Seconds  
(without alignments)  
1915.734 Million cell updates/sec

Title: US-09-807-802A-15

Perfect score: 3251  
Sequence: 1 TAPGKKRPVEQSPQEPDSSS.....NNGLYTEPRPIGTRVLTPL 599

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq 101002:\*

1: /SIDS2/gcgdata/geneseq/geneeqp-emb1/AA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneeqp-emb1/AA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneeqp-emb1/AA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneeqp-emb1/AA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneeqp-emb1/AA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneeqp-emb1/AA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneeqp-emb1/AA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneeqp-emb1/AA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneeqp-emb1/AA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneeqp-emb1/AA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneeqp-emb1/AA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneeqp-emb1/AA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneeqp-emb1/AA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneeqp-emb1/AA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneeqp-emb1/AA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneeqp-emb1/AA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneeqp-emb1/AA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneeqp-emb1/AA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneeqp-emb1/AA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneeqp-emb1/AA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneeqp-emb1/AA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneeqp-emb1/AA2001.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneeqp-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3251	100.0	599	21	AAV71168 Adeno-associated v
2	3251	100.0	736	21	AAV71167 Adeno-associated v
3	3229	99.3	736	22	AAB59847 AAV6 capsid protei
4	2906	89.4	534	21	AAV71169 Adeno-associated v
5	2832	87.1	736	22	AAB59846 AAV3B capsid prote
6	2815	86.6	736	22	AAB59845 AAV3A capsid prote
7	2764.5	85.0	735	22	AAG65792 Adeno-associated v
8	2764.5	85.0	735	22	AAM51508 Adeno-associated v
9	2764.5	85.0	735	22	AAB59844 AAV2 capsid protei
10	2759.5	84.9	598	22	AAG65793 Adeno-associated v

11	2759.5	84.9	598	22	AAM51509	Adeno-associated v
12	2738	84.2	734	22	AAB50326	Adeno-associated v
13	2481.5	76.3	533	22	AAG65794	Adeno-associated v
14	2477.5	76.2	533	22	AAM51510	Adeno-associated v
15	1830.5	56.3	598	19	AAW46313	AAV4 VP2 coat prot
16	1830.5	56.3	734	19	AAW46308	AAV4 VP1 capsid pr
17	1719	52.9	736	18	AAW19000	Duck parvovirus ca
18	1701	52.3	732	16	AAW85385	Barbary duck parvo
19	1700.5	52.3	588	21	AAV58161	Adeno associated v
20	1700.5	52.3	588	23	AAU11406	Adeno-associated v
21	1700.5	52.3	724	21	AAV58160	Adeno associated v
22	1700.5	52.3	724	23	AAU11405	Adeno-associated v
23	1690.5	52.0	544	19	AAW46314	AAV4 VP3 coat prot
24	1676	51.6	534	16	AAW85386	Barbary duck parvo
25	1665	51.2	532	21	AAV58162	Adeno associated v
26	1665	51.2	532	23	AAU11407	Adeno-associated v
27	772	23.7	756	21	AAV71231	Capsid protein enc
28	497.5	15.3	781	16	AAW08986	Human parvovirus V
29	497.5	15.3	781	20	AAV23227	Erythrovirus V9 VP
30	490	15.1	554	16	AAW08987	Human parvovirus V
31	490	15.1	554	20	AAV23230	Erythrovirus V9 VP
32	476.5	14.7	543	12	AAW13405	Parvo virus B19 VP
33	310	9.5	264	12	AAW13407	Parvo virus B19 PA
34	295.5	9.1	370	12	AAW13406	Parvo virus B19 PA
35	283	8.7	202	23	AAW83481	Adeno-associated v
36	250.5	7.7	686	5	AAW40068	Sequence of a porc
37	243.5	7.5	579	13	AAW29079	Porcine Parvovirus
38	216.5	6.7	579	14	AAW38697	PSY875 swine parvo
39	212.5	6.5	579	14	AAW38702	Swine parvovirus B
40	212.5	6.5	579	17	AAW99721	Swine parvovirus B
41	212.5	6.5	598	8	AAW70500	Pig parvo virus B
42	207.5	6.4	579	10	AAW94798	Swine parvovirus B
43	207	6.4	584	14	AAW30811	Feline parvovirus
44	203	6.2	472	5	AAW40675	Sequence encoded b
45	203	6.2	584	22	AAW04302	Feline panleukopen

ALIGNMENTS

RESULT 1	
AAV71168	Adeno-associated virus serotype 1; AAV-1; rep protein; capsid protein; cap protein; recombinant viral vector; gene delivery; gene therapy; vaccine; transgene; VP2.
AAV71168 standard; Protein; 599 AA.	
AAV71168;	
08-SEP-2000 (first entry)	
Adeno-associated virus serotype 1 capsid protein VP2.	
Adeno-associated virus serotype 1.	
WO200028061-A2.	
18-MAY-2000.	
02-NOV-1999; 99WO-US25694.	
05-NOV-1998; 98US-0107114.	
(TYPE-) UNIV PENNSYLVANIA.	
Wilson JM, Xiao W;	
WPI; 2000-376571/32.	
N-PSDB; AAD00772, AAD00778.	
Novel adeno-associated virus serotype 1 polynucleotide useful for preparation of medicament for delivery of a transgene to a host	



XX Claim 7; Page 93-95; 108bp; English.  
PS  
XX  
CC The patent discloses an adeno-associated virus serotype 1 (AAV-1) DNA  
CC which is characterised by two inverted terminal repeats (ITR) and open  
CC reading frames for rep and capsid (cap) proteins. The rep reading frame  
CC encodes four proteins, Rep 78, Rep 68, Rep 52 and Rep 40, while the cap  
CC reading frame encodes three structural proteins, VP1, VP2 and VP3.  
CC The AAV-1 sequence or its fragments particularly ITRs, rep and caps  
CC coding regions, are useful in production of recombinant viral vectors  
CC for gene delivery. These vectors can be used as gene therapy  
CC vectors, vaccine vectors or antisense delivery vectors. The AAV-1 does  
CC not induce the formation of neutralising antibodies specific to any  
CC serotype of AAV hence is useful for transforming host cells, and in the  
CC preparation of a medicament for the delivery of transgene to a host.  
CC The present sequence is an AAV-1 cap protein VP2 which is  
CC useful in the production of recombinant viral vector for gene delivery.

XX  
SQ Sequence 599 AA;

Query Match 100.0%; Score 3251; DB 21; Length 599;  
Best Local Similarity 100.0%; Pred. No. 1.3e-254;  
Matches 599; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAPGKRPEQSPQEPDSSSGIGKTGQOPAKKRLNFGQTGDSSESVDPDPLGEPATPAA 60  
DB 1 TAPGKRPEQSPQEPDSSSGIGKTGQOPAKKRLNFGQTGDSSESVDPDPLGEPATPAA 60  
QY 61 VGPTTMASGGGAPMADNNEGADGVGNASGNMHCSTWLGDRVITSTRTVALPTYNNHLY 120  
DB 61 VGPTTMASGGGAPMADNNEGADGVGNASGNMHCSTWLGDRVITSTRTVALPTYNNHLY 120  
QY 121 KOISSASTGASNDNHFGYSTPWCYFDENRPHCHFSPRDQRLINNNGFRPKRLNFKLF 180  
DB 121 KOISSASTGASNDNHFGYSTPWCYFDENRPHCHFSPRDQRLINNNGFRPKRLNFKLF 180  
QY 181 NIQVKEVTNDGVTTIANNLTSTVQVFSDEYQLPYVLGSAHQGLPPPADVFMIPOYG 240  
DB 181 NIQVKEVTNDGVTTIANNLTSTVQVFSDEYQLPYVLGSAHQGLPPPADVFMIPOYG 240  
QY 241 YLTINNGSQAVGRSSFYCLEYFPPSOMLRTGNFTFSYTFEEVPHSSYAHQSOLDRLMNP 300  
DB 241 YLTINNGSQAVGRSSFYCLEYFPPSOMLRTGNFTFSYTFEEVPHSSYAHQSOLDRLMNP 300  
QY 301 LIDQYLYLNRTQNGSGAQNKDLLFSRGSFAGMSVQPKNWLPGPCYRQORVSKTKTDNN 360  
DB 301 LIDQYLYLNRTQNGSGAQNKDLLFSRGSFAGMSVQPKNWLPGPCYRQORVSKTKTDNN 360  
QY 361 NSNFTWTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMSCVMI FGKESAGASNTALDN 420  
DB 361 NSNFTWTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMSCVMI FGKESAGASNTALDN 420  
QY 421 VMITDEEIKATNPVATERFGTVAVNFQSSSTDPATGVDHAMGALPGVWQDRDVLQGP 480  
DB 421 VMITDEEIKATNPVATERFGTVAVNFQSSSTDPATGVDHAMGALPGVWQDRDVLQGP 480  
QY 481 IWAKIPHTDGHFSPPLMGFGGLKNPPQILIKNTVPANPAAEFASATKFAFITQYSTG 540  
DB 481 IWAKIPHTDGHFSPPLMGFGGLKNPPQILIKNTVPANPAAEFASATKFAFITQYSTG 540  
QY 541 QVSVEIEWELOKENSKRWNPEVOYTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 599  
DB 541 QVSVEIEWELOKENSKRWNPEVOYTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 599

RESULT 2  
AAV71167  
ID AAV71167 standard; Protein; 736 AA.

XX AC AAV71167;  
XX DT 08-SEP-2000 (first entry)  
XX

DE Adeno-associated virus serotype 1 capsid protein VP1.  
XX  
XX Adeno-associated virus serotype 1; AAV-1; rep protein; capsid protein;  
KW cap protein; recombinant viral vector; gene delivery; gene therapy;  
KW vaccine; transgene; VP1.  
XX  
OS Adeno associated virus serotype 1.  
XX  
PN MO200028061-A2.  
XX  
PD 18-MAY-2000.  
XX  
PF 02-NOV-1999; 99WO-US25694.  
XX  
PR 05-NOV-1998; 98US-0107114.  
XX  
PA (TYPE-) UNIV PENNSYLVANIA.  
XX  
PI Wilson JM, Xiao W;  
XX  
DR WPI; 2000-376571/32.  
DR N-PSDB; AAD00772, AAD00777.  
XX  
XX

PT Novel adeno-associated virus serotype 1 polynucleotide useful for  
PT preparation of medicament for delivery of a transgene to a host  
XX  
PS Claim 7; Page 87-90; 108bp; English.

CC The patent discloses an adeno-associated virus serotype 1 (AAV-1) DNA  
CC which is characterised by two inverted terminal repeats (ITR) and open  
CC reading frames for rep and capsid (cap) proteins. The rep reading frame  
CC encodes four proteins, Rep 78, Rep 68, Rep 52 and Rep 40, while the cap  
CC reading frame encodes three structural proteins, VP1, VP2 and VP3.  
CC The AAV-1 sequence or its fragments particularly ITRs, rep and cap  
CC coding regions, are useful in production of recombinant viral vectors  
CC for gene delivery. These vectors can be used as gene therapy  
CC vectors, vaccine vectors or antisense delivery vectors. The AAV-1 does  
CC not induce the formation of neutralising antibodies specific to any  
CC serotype of AAV hence is useful for transforming host cells, and in the  
CC preparation of a medicament for the delivery of transgene to a host.  
CC The present sequence is an AAV-1 cap protein VP1 which is  
CC useful in the production of recombinant viral vector for gene delivery.

SQ Sequence 736 AA;

Query Match 100.0%; Score 3251; DB 21; Length 736;  
Best Local Similarity 100.0%; Pred. No. 1.7e-254;  
Matches 599; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAPGKRPEQSPQEPDSSSGIGKTGQOPAKKRLNFGQTGDSSESVDPDPLGEPATPAA 60  
DB 138 TAPGKRPEQSPQEPDSSSGIGKTGQOPAKKRLNFGQTGDSSESVDPDPLGEPATPAA 197  
QY 61 VGPTTMASGGGAPMADNNEGADGVGNASGNMHCSTWLGDRVITSTRTVALPTYNNHLY 120  
DB 198 VGPTTMASGGGAPMADNNEGADGVGNASGNMHCSTWLGDRVITSTRTVALPTYNNHLY 257  
QY 121 KOISSASTGASNDNHFGYSTPWCYFDENRPHCHFSPRDQRLINNNGFRPKRLNFKLF 180  
DB 258 KOISSASTGASNDNHFGYSTPWCYFDENRPHCHFSPRDQRLINNNGFRPKRLNFKLF 317  
QY 181 NIQVKEVTNDGVTTIANNLTSTVQVFSDEYQLPYVLGSAHQGLPPPADVFMIPOYG 240  
DB 318 NIQVKEVTNDGVTTIANNLTSTVQVFSDEYQLPYVLGSAHQGLPPPADVFMIPOYG 377  
QY 241 YLTINNGSQAVGRSSFYCLEYFPPSOMLRTGNFTFSYTFEEVPHSSYAHQSOLDRLMNP 300  
DB 378 YLTINNGSQAVGRSSFYCLEYFPPSOMLRTGNFTFSYTFEEVPHSSYAHQSOLDRLMNP 437  
QY 301 LIDQYLYLNRTQNGSGAQNKDLLFSRGSFAGMSVQPKNWLPGPCYRQORVSKTKTDNN 360  
DB 438 LIDQYLYLNRTQNGSGAQNKDLLFSRGSFAGMSVQPKNWLPGPCYRQORVSKTKTDNN 497

OY 361 NSNFTWTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMGVMIFGKESAGASNTALDN 420  
DB 498 NSNFTWTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMGVMIFGKESAGASNTALDN 557  
OY 421 VMITDEEEIKATNPVATERFGTVAVNFQSSSTDPAATGDVHAMGALPGMWQDRDVLQGP 480  
DB 558 VMITDEEEIKATNPVATERFGTVAVNFQSSSTDPAATGDVHAMGALPGMWQDRDVLQGP 617  
OY 481 IWAKIPHTDGHFHPSPILMGFGFLKNPPQILIKNTVPANPAEFSATKFAFITQYSTG 540  
DB 618 IWAKIPHTDGHFHPSPILMGFGFLKNPPQILIKNTVPANPAEFSATKFAFITQYSTG 677  
OY 541 QVSVEIEMELQENSKRMNPEVOYTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 599  
DB 678 QVSVEIEMELQENSKRMNPEVOYTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 736

RESULT 3  
AAB59847

ID AAB59847 standard; Protein; 736 AA.

AC AAB59847;

DT 28-MAR-2001 (first entry)

DE AAV6 capsid protein VP1.

KW AAV6; gene therapy; AAV viral vector; cystic fibrosis; cancer; AIDS;

KW atherosclerosis; sickle cell anaemia; thalassaemia;

KW blood clotting disorder; diabetes; capsid protein VP1.

OS Adeno associated virus.

PN US6156303-A.

PD 05-DEC-2000.

PF 11-JUN-1997; 97US-0873168.

PR 11-JUN-1997; 97US-0873168.

PA (UNIW ) UNIV WASHINGTON.

PI Russell DW, Rutledge EA;

DR WPI; 2001-060164/07.

XX  
PT Adeno-associated virus serotype 6 and viral vector derived from it for  
PT gene therapy of cystic fibrosis, cancer, acquired immunodeficiency  
PT syndrome, sickle cell anemia, thalassemia and diabetes  
XX  
PS Disclosure; Fig 2; 50pp; English.  
XX  
CC The present invention relates to adeno-associated virus serotypes. The  
CC present sequence is capsid protein VP1 of one such serotype (AAV6). AAV6  
CC can be used to construct AAV viral vectors for use in gene therapy for a  
CC range of disorders: cystic fibrosis, cancer, AIDS, atherosclerosis,  
CC sickle cell anaemia, thalassaemia, blood clotting disorders and diabetes.  
CC The AAV viral vectors have increased transduction efficiency of a  
CC particular host cell as the AAV virion containing the AAV vector genome  
CC can be modified to express a capsid protein of an AAV serotype that  
CC transduces the selected host cell.  
XX

XX  
SQ Sequence 736 AA;

Query Match 99.3%; Score 3229; DB 22; Length 736; .

Best Local Similarity 99.2%; Pred. No. 1e-252;

Matches 594; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 1 TAGGKKRPVEQSPQEPDSSSGIGKTGQPAKKRLNFGQTGDSSESVDPDPLGEPATPAA 60  
DB 138 TAGGKKRPVEQSPQEPDSSSGIGKTGQPAKKRLNFGQTGDSSESVDPDPLGEPATPAA 197

OY 61 VGPTTASGGGAPMADNNEGADGVNAGSNWCHDSTWLGDRIITSTRTWALPTVNNHLY 120  
DB 198 VGPTTASGGGAPMADNNEGADGVNAGSNWCHDSTWLGDRIITSTRTWALPTVNNHLY 257  
OY 121 KOISSASTGASNDNHFGYSTPWGYFEDENRFHCHFSPRDQRLINNMGFRPKRLNFKLF 180  
DB 258 KOISSASTGASNDNHFGYSTPWGYFEDENRFHCHFSPRDQRLINNMGFRPKRLNFKLF 317  
OY 181 NIQVEVTTNDGVTTIANNLSTVQVESDSEYQLPYVLGSAHQCLPPPADVFMIPOYG 240  
DB 318 NIQVEVTTNDGVTTIANNLSTVQVESDSEYQLPYVLGSAHQCLPPPADVFMIPOYG 377  
OY 241 YLTINNGSQAVGRSSFYCLEYFPSSQMLRTGNFTFSYTFEEVFPHSSYAHSSQSLDRMNP 300  
DB 378 YLTINNGSQAVGRSSFYCLEYFPSSQMLRTGNFTFSYTFEEVFPHSSYAHSSQSLDRMNP 437  
OY 301 LIDQYLYLNRTONQSGAQNKDLLFSRGS PAGMSVQPKWLPGPCYRQORVSKTIDNN 360  
DB 438 LIDQYLYLNRTONQSGAQNKDLLFSRGS PAGMSVQPKWLPGPCYRQORVSKTIDNN 497  
OY 361 NSNFTWTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMGVMIFGKESAGASNTALDN 420  
DB 498 NSNFTWTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMGVMIFGKESAGASNTALDN 557  
OY 421 VMITDEEEIKATNPVATERFGTVAVNFQSSSTDPAATGDVHAMGALPGMWQDRDVLQGP 480  
DB 558 VMITDEEEIKATNPVATERFGTVAVNLQSSSTDPAATGDVHAMGALPGMWQDRDVLQGP 617  
OY 481 IWAKIPHTDGHFHPSPILMGFGFLKNPPQILIKNTVPANPAEFSATKFAFITQYSTG 540  
DB 618 IWAKIPHTDGHFHPSPILMGFGFLKNPPQILIKNTVPANPAEFSATKFAFITQYSTG 677  
OY 541 QVSVEIEMELQENSKRMNPEVOYTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 599  
DB 678 QVSVEIEMELQENSKRMNPEVOYTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 736

RESULT 4  
AA71169

ID AA71169 standard; Protein; 534 AA.

AC AA71169;

DT 08-SEP-2000 (first entry)

DE Adeno-associated virus serotype 1 capsid protein VP3.

KW Adeno-associated virus serotype 1; AAV-1; rep protein; capsid protein;

KW cap protein; recombinant viral vector; gene delivery; gene therapy;

KW vaccine; transgene; VP3.

OS Adeno associated virus serotype 1.

PN WO200028061-A2.

PD 18-MAY-2000.

PF 02-NOV-1999; 99WO-US25694.

PR 05-NOV-1998; 98US-0107114.

PA (UYPE-) UNIV PENNSYLVANIA.

PI Wilson JM, Xiao W;

DR WPI; 2000-376571/32.

DR N-PSDB; AAD00772, AAD00779.

PT Novel adeno-associated virus serotype 1 polynucleotide useful for  
PT preparation of medicament for delivery of a transgene to a host  
XX  
PS Claim 7; Page 99-101; 108pp; English.  
XX

CC The patent discloses an adeno-associated virus serotype 1 (AAV-1) DNA  
CC which is characterised by two inverted terminal repeats (ITR) and open  
CC reading frames for rep and capsid (cap) proteins. The rep reading frame  
CC encodes four proteins, Rep 78, Rep 68, Rep 52 and Rep 40, while the cap  
CC reading frame encodes three structural proteins, VP1, VP2 and VP3.  
CC The AAV-1 sequence or its fragments particularly ITRs, rep and cap  
CC coding regions, are useful in production of recombinant viral vectors  
CC for gene delivery. These vectors can be used as gene therapy  
CC vectors, vaccine vectors or antisense delivery vectors. The AAV-1 does  
CC not induce the formation of neutralising antibodies specific to any  
CC serotype of AAV hence is useful for transforming host cells, and in the  
CC preparation of a medicament for the delivery of transgene to a host.  
CC The present sequence is an AAV-1 cap protein VP3 which is  
CC useful in the production of recombinant viral vector for gene delivery.

XX Sequence 534 AA;

Query Match 89.4%; Score 2906; DB 21; Length 534;  
Best Local Similarity 100.0%; Pred. No. 9.7e-227;  
Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 MASGGAPMADNNEGADGVGNASGNMHCDSWLGDVITTTSTRTWALPTNNHLYKQISS 125  
DB 1 MASGGAPMADNNEGADGVGNASGNMHCDSWLGDVITTTSTRTWALPTNNHLYKQISS 60  
QY 126 ASTGASNDNHFGYSTPWGYPDFNRFHCHFSRDMQRLINNMWGFPRKRLNFKLFNIQVK 185  
DB 61 ASTGASNDNHFGYSTPWGYPDFNRFHCHFSRDMQRLINNMWGFPRKRLNFKLFNIQVK 120  
QY 186 EVTTNDGVTTIANNLTSTVQVFSSEYQLPYVLGSAHQGCLPPFPADVFMIPQYGYLTIN 245  
DB 121 EVTTNDGVTTIANNLTSTVQVFSSEYQLPYVLGSAHQGCLPPFPADVFMIPQYGYLTIN 180  
QY 246 NGSQAVGRSSFYCLEYFPPSQMLRTGNFTFSYTFEEVPEFHSYAHQSQSLDRMLNPLIDQY 305  
DB 181 NGSQAVGRSSFYCLEYFPPSQMLRTGNFTFSYTFEEVPEFHSYAHQSQSLDRMLNPLIDQY 240  
QY 306 LYYLNRTQNGSGAQNKDLLFSRGSPPAGMSVQPKMWLPGPCYRQQRVSKTKTDNNNSNFT 365  
DB 241 LYYLNRTQNGSGAQNKDLLFSRGSPPAGMSVQPKMWLPGPCYRQQRVSKTKTDNNNSNFT 300  
QY 366 WTGASKYNLNGRESINPGTAMASHKDEDEKFFPMSCVMI FGKESAGASNTALDNVMITD 425  
DB 301 WTGASKYNLNGRESINPGTAMASHKDEDEKFFPMSCVMI FGKESAGASNTALDNVMITD 360  
QY 426 EEIKATNPVATERFGTVAVNFQSSSTDPATGVDHAMGALPGMWQDRDVIYLGPIWAKI 485  
DB 361 EEIKATNPVATERFGTVAVNFQSSSTDPATGVDHAMGALPGMWQDRDVIYLGPIWAKI 420  
QY 486 PHTDGHFHPSPLMGFGGLKNPPQIILIKNTVPVNPAPAEFSATKFASTIQYSTGQSVSE 545  
DB 421 PHTDGHFHPSPLMGFGGLKNPPQIILIKNTVPVNPAPAEFSATKFASTIQYSTGQSVSE 480  
QY 546 IEMELQKENSKRWPEVOYTSNYAKSANVDFVDNNGLYTEPRPIGTRYLTRPL 599  
DB 481 IEMELQKENSKRWPEVOYTSNYAKSANVDFVDNNGLYTEPRPIGTRYLTRPL 534

RESULT 5  
AAB59846 AAB59846 standard; Protein; 736 AA.  
XX AC AAB59846;  
XX DT 28-MAR-2001 (first entry)  
XX DE AAV3B capsid protein VP1.  
XX KM AAV3B; gene therapy; AAV viral vector; cystic fibrosis; cancer; AIDS;  
XX KM atherosclerosis; sickle cell anaemia; thalassaemia;  
XX KM blood clotting disorder; diabetes; capsid protein VP1.  
OS Adeno associated virus.

XX PN US6156303-A.  
XX PD 05-DEC-2000.  
XX PF 11-JUN-1997; 97US-0873168.  
XX PR 11-JUN-1997; 97US-0873168.  
XX PA (UNIW ) UNIV WASHINGTON.  
XX PI Russell DW, Rutledge EA;  
XX DR WPI; 2001-060164/07.

PT Adeno-associated virus serotype 6 and viral vector derived from it for  
PT gene therapy of cystic fibrosis, cancer, acquired immunodeficiency  
PT syndrome, sickle cell anemia, thalassemia and diabetes  
XX PS Disclosure; Fig 2; 50pp; English.

CC The present invention relates to adeno-associated virus serotypes. The  
CC present sequence is capsid protein VP1 of one such serotype (AAV3B).  
CC AAV3B can be used to construct AAV viral vectors for use in gene therapy  
CC for a range of disorders: cystic fibrosis, cancer, AIDS, atherosclerosis,  
CC sickle cell anaemia, thalassaemia, blood clotting disorders and diabetes.  
CC The AAV viral vectors have increased transduction efficiency of a  
CC particular host cell as the AAV virion containing the AAV vector genome  
CC can be modified to express a capsid protein of an AAV serotype that  
CC transduces the selected host cell.

SQ Sequence 736 AA;

Query Match 87.1%; Score 2832; DB 22; Length 736;  
Best Local Similarity 85.7%; Pred. No. 1.5e-220;  
Matches 514; Conservative 34; Mismatches 50; Indels 2; Gaps 2;

QY 1 TAPGKKRPVEQSPOEPDSSSGIGTKTGOQPAKKRLNFGQTDSESVDPDQPLGEPATPAA 60  
DB 138 TAPGKKRPVDQSPQEPDSSSGVGKSGKQPAKKRLNFGQTDSESVDPDQPLGEPATPAA 197  
QY 61 VGPPTMASGGGAPMADNNEGADGVGNASGNMHCDSWLGDVITTTSTRTWALPTNNHLY 120  
DB 198 LGSNTMASGGGAPMADNNEGADGVGNASGNMHCDSWLGDVITTTSTRTWALPTNNHLY 257  
QY 121 KOISSASTGASNDNHFGYSTPWGYPDFNRFHCHFSRDMQRLINNMWGFPRKRLNFKLF 180  
DB 258 KOISSQS-GASNDNHFGYSTPWGYPDFNRFHCHFSRDMQRLINNMWGFPRKRLNFKLF 316  
QY 181 NIOVKEVTNDGVTTIANNLTSTVQVFSSEYQLPYVLGSAHQGCLPPFPADVFMIPQY 240  
DB 317 NIOVKEVTNDGVTTIANNLTSTVQVFTDSEYQLPYVLGSAHQGCLPPFPADVFMIPQY 376  
QY 241 YLTNNGSOAVGRSSFYCLEYFPPSQMLRTGNFTFSYTFEEVPEFHSYAHQSQSLDRMLNP 300  
DB 377 YLTNNGSOAVGRSSFYCLEYFPPSQMLRTGNFTFSYTFEEVPEFHSYAHQSQSLDRMLNP 436  
QY 301 LIDQVLYLNRTQ-NQSGAQNKDLLFSRGSPPAGMSVQPKMWLPGPCYRQQRVSKTKTDN 359  
DB 437 LIDQVLYLNRTQGTSGTNGRLLFSQAGPQMSLQARNWLPGPCYRQQRVSKTKTDN 496  
QY 360 NNSNFTWTGASKYNLNGRESINPGTAMASHKDEDEKFFPMSCVMI FGKESAGASNTALD 419  
DB 497 NNSNFTWTGASKYNLNGRDSLVPDPAMASHKDEDEKFFPMHGNLIFGKEGTTASNAELD 556  
QY 420 NVMITDEEIKATNPVATERFGTVAVNFQSSSTDPATGVDHAMGALPGMWQDRDVIYLG 479  
DB 557 NVMITDEEIKATNPVATERFGTVAVNFQSSSTDPATGVDHAMGALPGMWQDRDVIYLG 616  
QY 480 PIWAKIPHTDGHFHPSPLMGFGGLKNPPQIILIKNTVPVNPAPAEFSATKFASTIQYST 539  
DB 617 PIWAKIPHTDGHFHPSPLMGFGGLKNPPQIILIKNTVPVNPAPAEFSATKFASTIQYST 676





CC the cell surface and permits the vector to enter the cell efficiently.  
CC A pharmaceutical composition comprising RRV with (i) containing a  
CC transgene sequence associated with a disease or a disorder such that  
CC expression of the transgene would result in amelioration of the disease  
CC or disorder such as inherited neurological and metabolic diseases e.g.  
CC lysosomal storage disease, Lesch-Nyhan syndrome, amyloid polynuropathy,  
CC Alzheimer amyloid, Duchenne's muscular dystrophy, diseases of the blood  
CC e.g. sickle-cell anemia, clotting disorders and thalassemias, cystic  
CC fibrosis, diabetes, diseases associated with hormone deficiencies,  
CC retinoblastoma and various types of neoplastic cells which include  
CC tumours especially central nervous system tumours, neoplasms, carcinomas,  
CC sarcomas, leukemias and lymphoma. The present sequence represents the  
CC adeno-associated virus 2 (AAV-2) major coat protein VP1. AAV-2 sequences  
CC are used in the construction of a chimeric vector.

XX Sequence 735 AA;

Query Match 85.0%; Score 2764.5; DB 22; Length 735;  
Best Local Similarity 82.8%; Pred. No. 4.4e-215;  
Matches 496; Conservative 43; Mismatches 59; Indels 1; Gaps 1;

QY 1 TAPGKRPEVQSPQEPDSSSGIGKTGOQPAKRLNFGQTGDSSEVPDPOPLGEPPATPAA 60  
DB 138 TAPGKRPEVHSPVEPDSSSGTGAKGOQPAKRLNFGQTGDADSVDPDPOPLGPPAPPSG 197  
QY 61 VGPTMASGGAPMADNNEGADGVGNASGNWCHDSTWLGDRVITTSRTWALPTYNNHLY 120  
DB 198 LGTNTMATSGAPMADNNEGADGVGNSSGNWCHDSTWMDRVITTSRTWALPTYNNHLY 257  
QY 121 KOISSASTGASNDNHFGYSTPWGYFDENRFHCHFSRDMQRLINNNGFRPKRLNFKLF 180  
DB 258 KOISSQS-GASNDNHFGYSTPWGYFDENRFHCHFSRDMQRLINNNGFRPKRLNFKLF 316  
QY 181 NIQVKEVTNDGVTTIANNLTSTVQVFSDESEYQLPYVLGSAHQGLPPPADVFMIPQYG 240  
DB 317 NIQVKEVTQNDGTTIANNLSTVQVFTDSEYQLPYVLGSAHQGLPPPADVFMVPOYG 376  
QY 241 YLTLNNGSQAVGRSSFYCLEYFSPQMLRTGNNTFSYTFEEVPHSSSYAHSQSLDRLMNP 300  
DB 377 YLTLNNGSQAVGRSSFYCLEYFSPQMLRTGNNTFSYTFEDVPHSSSYAHSQSLDRLMNP 436  
QY 301 LIDQLYLYLNRTQNGSGSAQNKDLLFSRGSPPAGMSVQPKNMLPGPCYRQORVSKTDTNN 360  
DB 437 LIDQLYLYLSRTNTPSGTTTQSRLOFSQAGASDIRDQSRNMLPGPCYRQORVSKTSADNN 496  
QY 361 NSNFTWTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMGVMIFGKESAGASNTALDN 420  
DB 497 NSEYSWTGATKYHLNGRDSLVPNGPAMASHKDEDEKFFPQSGVLI FGKQSEKTNVDIEK 556  
QY 421 VMITDEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMGALPGMWQDRDVLQGP 480  
DB 557 VMITDEEIRITNPVATEQYGSVSTNLQGRNQATADVNTQGVLPGMWQDRDVLQGP 616  
QY 481 IWAKIPHTDGHFHPSPMLMGFGGLKNPPOILIKNTVPANPPAEFSATKPFASFITQYSTG 540  
DB 617 IWAKIPHTDGHFHPSPMLMGFGGLKHPPOILIKNTVPANPSTTFSAKPFASFITQYSTG 676  
QY 541 QVSVEIEMWELQKENS KRWNPVEVQYTSNYAKSANVDFTVNNGLYTEPRPIGTRYLTRPL 599  
DB 677 QVSVEIEMWELQKENS KRWNPFIQYTSNYNKSANVDFTVDTNGVYSEPRPIGTRYLTRNL 735

RESULT 8

AAAM51508  
ID AAM51508 standard; protein; 735 AA.

XX AC AAM51508;

DT 02-JAN-2002 (first entry)

XX Adeno-associated virus VP1 capsid protein.

XX Adeno-associated virus; AAV; VP1; capsid; virus-like particle;

KW nuclear localisation signal; VP3.

XX Adeno associated virus.

XX JP2001169777-A.

XX 26-JUN-2001.

XX 30-JUL-1999; 99JP-0249140.

XX 30-JUL-1999; 99JP-0249140.

XX (HAND/) HANDA H.

XX MPI; 2001-599854/68.

PT New virus-like particles from VP3 capsid protein of adeno-associated  
PT virus, comprise a peptide containing a nucleus-shifting signal  
PT connected to its N-terminal -

PS Disclosure; Page 10-13; 33pp; Japanese.

CC The present sequence is provided in a specification relating to a  
CC virus-like particle-forming protein, and to a peptide containing a  
CC nuclear-shifting signal at its N-terminus that can form virus-like  
CC particles by shifting to the nucleus of the animal cell in which it  
CC is expressed. The method is used for forming virus-like particles  
CC from the VP3 capsid protein of adeno-associated virus (AAV).

XX Sequence 735 AA;

Query Match 85.0%; Score 2764.5; DB 22; Length 735;  
Best Local Similarity 82.8%; Pred. No. 4.4e-215;  
Matches 496; Conservative 43; Mismatches 59; Indels 1; Gaps 1;

QY 1 TAPGKRPEVQSPQEPDSSSGIGKTGOQPAKRLNFGQTGDSSEVPDPOPLGEPPATPAA 60  
DB 138 TAPGKRPEVHSPVEPDSSSGTGAKGOQPAKRLNFGQTGDADSVDPDPOPLGPPAPPSG 197  
QY 61 VGPTMASGGAPMADNNEGADGVGNASGNWCHDSTWLGDRVITTSRTWALPTYNNHLY 120  
DB 198 LGTNTMATSGAPMADNNEGADGVGNSSGNWCHDSTWMDRVITTSRTWALPTYNNHLY 257  
QY 121 KOISSASTGASNDNHFGYSTPWGYFDENRFHCHFSRDMQRLINNNGFRPKRLNFKLF 180  
DB 258 KOISSQS-GASNDNHFGYSTPWGYFDENRFHCHFSRDMQRLINNNGFRPKRLNFKLF 316  
QY 181 NIQVKEVTNDGVTTIANNLTSTVQVFSDESEYQLPYVLGSAHQGLPPPADVFMIPQYG 240  
DB 317 NIQVKEVTQNDGTTIANNLSTVQVFTDSEYQLPYVLGSAHQGLPPPADVFMVPOYG 376  
QY 241 YLTLNNGSQAVGRSSFYCLEYFSPQMLRTGNNTFSYTFEEVPHSSSYAHSQSLDRLMNP 300  
DB 377 YLTLNNGSQAVGRSSFYCLEYFSPQMLRTGNNTFSYTFEDVPHSSSYAHSQSLDRLMNP 436  
QY 301 LIDQLYLYLNRTQNGSGSAQNKDLLFSRGSPPAGMSVQPKNMLPGPCYRQORVSKTDTNN 360  
DB 437 LIDQLYLYLSRTNTPSGTTTQSRLOFSQAGASDIRDQSRNMLPGPCYRQORVSKTSADNN 496  
QY 481 IWAKIPHTDGHFHPSPMLMGFGGLKNPPOILIKNTVPANPPAEFSATKPFASFITQYSTG 540  
DB 617 IWAKIPHTDGHFHPSPMLMGFGGLKHPPOILIKNTVPANPSTTFSAKPFASFITQYSTG 676  
QY 541 QVSVEIEMWELQKENS KRWNPVEVQYTSNYAKSANVDFTVNNGLYTEPRPIGTRYLTRPL 599  
DB 677 QVSVEIEMWELQKENS KRWNPFIQYTSNYNKSANVDFTVDTNGVYSEPRPIGTRYLTRNL 735

Db 677 QVSVEIEMELQKENS KRWNP EIQYTSN YKSVN VDFTVDTNGVYSEPRPIGTRYLTRNL 735

RESULT 9  
AAB59844  
ID AAB59844 standard; Protein; 735 AA.

XX  
AC AAB59844;  
XX  
DT 28-MAR-2001 (first entry)  
XX  
DE AAV2 capsid protein VP1.  
XX  
KW AAV2; gene therapy; AAV viral vector; cystic fibrosis; cancer; AIDS;  
KW atherosclerosis; sickle cell anaemia; thalassaemia;  
KW blood clotting disorder; diabetes; capsid protein VP1.  
XX  
OS Adeno associated virus.  
XX  
PN US6156303-A.  
XX  
PD 05-DEC-2000.  
XX  
PF 11-JUN-1997; 97US-0873168.  
XX  
PR 11-JUN-1997; 97US-0873168.  
XX  
PA (UNIW ) UNIV WASHINGTON.  
XX  
PI Russell DW, Rutledge EA;  
XX  
DR WPI; 2001-060164/07.

XX  
PT Adeno-associated virus serotype 6 and viral vector derived from it for  
PT gene therapy of cystic fibrosis, cancer, acquired immunodeficiency  
PT syndrome, sickle cell anemia, thalassemia and diabetes -  
XX  
XX  
PS Claim 7; Fig 2; 50pp; English.

XX  
CC The present invention relates to adeno-associated virus serotypes. The  
CC present sequence is capsid protein VP1 of one such serotype (AAV2). AAV2  
CC can be used to construct AAV viral vectors for use in gene therapy for a  
CC range of disorders: cystic fibrosis, cancer, AIDS, atherosclerosis,  
CC sickle cell anaemia, thalassaemia, blood clotting disorders and diabetes.  
CC The AAV viral vectors have increased transduction efficiency of a  
CC particular host cell as the AAV virion containing the AAV vector genome  
CC can be modified to express a capsid protein of an AAV serotype that  
CC transduces the selected host cell.  
XX  
SQ Sequence 735 AA;

Query Match 85.0%; Score 2764.5; DB 22; Length 735;  
Best Local Similarity 82.8%; Pred. No. 4.4e-215;  
Matches 496; Conservative 43; Mismatches 59; Indels 1; Gaps 1;

QY 1 TAPGKKRPVEQSPQEPDSSSGIGTKTGQOPAKKRLNFGQTGDSSEVPDPQPLGEPATPA 60  
DB 138 TAPGKKRPVEHSPVEPDDSSSGTGKAGQOPAKKRLNFGQTGDADSVDPQPLGQPPAAPSG 197

QY 61 VGPPTMASGGGAPMADNNEGADGVNAGNWHCDSTWLGDRVITTTSTRTVALPTYNHLY 120  
DB 198 LGTNTMA TGSGAPMADNNEGADGVNAGNWHCDSTWMDRVITTTSTRTVALPTYNHLY 257

QY 121 KQISSASTGASNDNHFGYSTPWPGYFDENRFCHFSPRDWQRLINNNGFRPKRLNFKLF 180  
DB 258 KQISSQG-GASNDNHFGYSTPWPGYFDENRFCHFSPRDWQRLINNNGFRPKRLNFKLF 316

QY 181 NIQVKEVTNDGVTTIANNLTSTVQVPSDSEYQLPYVLGSAHOGCLPPFPADVFMIPQYG 240  
DB 317 NIQVKEVTQNDGVTTIANNLTSTVQVFTDSEYQLPYVLGSAHOGCLPPFPADVFMVPOYG 376

QY 241 YLTLNNGSOAVGRSSFCLEYFSPQMLRTGNNTFTSFEEVFPFHSSYAHQSGLDRLMNP 300

Db 377 YLTLNNGSOAVGRSSFCLEYFSPQMLRTGNNTFTSFEEVFPFHSSYAHQSGLDRLMNP 436

QY 301 LIDQYLYLINRTQNGSAQNKDLLEFSRGS PAGMSVOPKNWLPGPCYRQORVSKTKTDNN 360  
DB 437 LIDQYLYLSTRNTPSGTTQSRLOFSQAGASDIRDQSRNWLPGPCYRQORVSKTSADNN 496

QY 361 NSNFTWTGASKYNLNGRESIINPGTAMASHKDEDEKFPFMSGMI FGKESAGASNTALDN 420  
DB 497 NSEYSWTGATKYHLNGRDSLVPNGPAMASHKDEDEKFPFQSGVLIFGKQSEKTNVDIEK 556

QY 421 VMITDEEIKATNPVATERFGTVAVNFOSSSTDPATGDVHAMGALPGMWODRDVYLOGP 480  
DB 557 VMITDEEIRTPNPVATEQYGSYSTNLORGNRQAATADVNTQGVLPGMWODRDVYLOGP 616

QY 481 IWAKIPHTDGHFHPSPLMGGFGLKNPPQIILKNTPVANPPAEFSATKFASFITQYSTG 540  
DB 617 IWAKIPHTDGHFHPSPLMGGFGLKHPPOILLKNTPVANPSTTFSAKFAFIFTQYSTG 676

QY 541 QVSVEIEMELQKENS KRWNP EIQYTSN YKSAVNDFTVNNGLYTEPRPIGTRYLTRPL 599  
DB 677 QVSVEIEMELQKENS KRWNP EIQYTSN YKSVN VDFTVDTNGVYSEPRPIGTRYLTRNL 735

RESULT 10  
AAG65793  
ID AAG65793 standard; Protein; 598 AA.  
XX  
AC AAG65793;  
XX  
DT 11-FEB-2002 (first entry)  
XX  
DE Adeno-associated virus 2 (AAV-2) major coat protein VP2.  
XX  
KW Recombinant viral vector; RVV; capsid; parvovirus; transgene; cytostatic;  
KW inverted terminal repeat; nootropic; neuroprotective; antianemic; ITR;  
KW antidiabetic; antitumour; gene therapy; adeno-associated virus; AAV;  
KW major coat protein; AAV-2; VP2.  
XX  
OS Adeno-associated virus 2.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 1 /note= "encoded by ACG"  
FT  
PN WO200168888-A2.  
XX  
PD 20-SEP-2001.  
XX  
PF 13-MAR-2001; 2001WO-US07927.  
XX  
PR 14-MAR-2000; 2000US-189110P.  
XX  
PA (NEUR-) NEUROLOGIX INC.  
XX  
PI Xiao W, During MJ;  
XX  
DR WPI; 2001-596912/67.  
XX  
N-PSDB; AA166974.

PT Recombinant viral vector useful in improving gene therapy in a subject,  
PT and for increasing efficiency of entry into a cell, comprises a  
PT chimeric capsid having one non-native amino acid sequence and a desired  
PT transgene -  
XX  
PS Disclosure; Page 51; 53pp; English.  
XX  
CC The invention provides a recombinant viral vector (RVV) comprising a  
CC chimeric capsid (I) having at least one non-native amino acid sequence,  
CC derived from a capsid protein domain of parvovirus (II), a virus (III),  
CC or their combination, and a transgene flanked 5' and 3' by inverted  
CC terminal repeat (ITR) sequences, derived from (II), (III), or their  
CC combination. The RVV is useful for improving gene therapy in a subject  
CC with a disorder, and for increasing the efficiency of entry into a cell,



CC which involves producing (I) encapsulating a viral vector, and contacting  
CC a cell with RVV having (I) such that (I) binds to an attachment site on  
CC the cell surface and permits the vector to enter the cell efficiently.  
CC A pharmaceutical composition comprising RVV with (I) containing a  
CC transgene sequence associated with a disease or a disorder such that  
CC expression of the transgene would result in amelioration of the disease  
CC or disorder such as inherited neurological and metabolic diseases e.g.  
CC lysosomal storage disease, Lesch-Nyhan syndrome, amyloid polynuropathy,  
CC Alzheimer amyloid, Duchenne's muscular dystrophy, diseases of the blood  
CC e.g. sickle-cell anemia, clotting disorders and thalassemias, cystic  
CC fibrosis, diabetes, diseases associated with hormone deficiencies,  
CC retinoblastoma and various types of neoplastic cells which include  
CC tumours especially central nervous system tumours, neoplasms, carcinomas,  
CC sarcomas, leukemias and lymphoma. The present sequence represents the  
CC adeno-associated virus 2 (AAV-2) major coat protein VP2. AAV-2 sequences  
CC are used in the construction of a chimeric vector.

XX Sequence 598 AA;

Query Match 84.9%; Score 2759.5; DB 22; Length 598;  
Best Local Similarity 82.8%; Pred. No. 8.4e-215;  
Matches 495; Conservative 43; Mismatches 59; Indels 1; Gaps 1;

QY 2 APGKKRPEVQSPQEPDSSSGIGKTGQOPAKKRLNFGQTDSESVDPDPLGEPATPAV 61  
DB 2 APGKKRPEVHSPVEPDSSSGTGAGQOPARKRLNFGQTDADSVDPDPLGPPAPPSGL 61  
QY 62 GPTTMASGGAPMADNNEGADGVGNASGNWCHDSTWLGDVITTTSTRTVALPTYNHLYK 121  
DB 62 GTNTMATGSGAPMADNNEGADGVGNSSGNWCHDSTWMDRVITTTSTRTVALPTYNHLYK 121  
QY 122 QISSASTGASNDNHFGYSTPWGYFDNFRFHCHFSPRDWORLNNNWGFRPKRLNFKLFN 181  
DB 122 QISSQS-GASNDNHFGYSTPWGYFDNFRFHCHFSPRDWORLNNNWGFRPKRLNFKLFN 180  
QY 182 IQVKEVTNDGVTIANNLSTVQVFSDESYQLPYVLGSAHQGLPPFPADVFMIPQYGY 241  
DB 181 IQVKEVTQNDGTTIANNLSTVQVFTDSEYQLPYVLGSAHQGLPPFPADVFMVPOYGY 240  
QY 242 LTLNNGSQAVGRSSFYCLEYFSPQMLRTGNNFTFSYTFEEVPHSSYAHQSOLDRLMNP 301  
DB 241 LTLNNGSQAVGRSSFYCLEYFSPQMLRTGNNFTFSYTFEDVFFHSSYAHQSOLDRLMNP 300  
QY 302 IDQYLYLNRTONQSGSAQNKDLLFSRGSFAGMSVQPKNMLPGPCYRQORVSKTKTDNN 361  
DB 301 IDQYLYLNRTONQSGTTTQSRLOFSQAGASDIRDQSRNMLPGPCYRQORVSKTSADNN 360  
QY 362 SNFTWTGASKYNLNGRESINPGTAMASHKDEDEKFFPMGVMIFGKESAGASNTALDNV 421  
DB 361 SEYSWTGATKYHLNGRDSLVPNPAMASHKDEKFFPQSGVLIFFKQOSEKTNVDIEKV 420  
QY 422 MITDEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMGALPGMWQDRDVLQGP 481  
DB 421 MITDEEIRTTNPVATEQYGSVSTNLQGRNQAATADVNTQGVLPGMWQDRDVLQGP 480  
QY 482 WAKIPHTDGHFHPSPMLMGFGLKNPPQILIKNTPVPANPFAFSATKFAFITQYSTGQ 541  
DB 481 WAKIPHTDGHFHPSPMLMGFGLKHPPQILIKNTPVPANPSTTFSAAKFAFITQYSTGQ 540  
QY 542 VSVEIEMELQKENSKRMPREVQTSNYAKSANVDFTVDNNGI.YTEPRPTGTRYLTRPL 599  
DB 541 VSVEIEMELQKENSKRMPREIOYTSNYNKSANVDFTVDNGVYSEPRPIGTRYLTRNL 598

RESULT 11

AAM51509 standard; protein; 598 AA.

AC AAM51509;

DT 02-JAN-2002 (first entry)

DE Adeno-associated virus VP2 capsid protein.

XX Adeno-associated virus; AAV; VP2; capsid; virus-like particle;  
KW nuclear localisation signal; VP3.  
XX Adeno associated virus.  
OS JP2001169777-A.  
XX 26-JUN-2001.  
XX 30-JUL-1999; 99JP-0249140.  
XX 30-JUL-1999; 99JP-0249140.  
XX 30-JUL-1999; 99JP-0249140.  
XX (HANA/) HANA H.  
XX WPI; 2001-599854/68.

PT New virus-like particles from VP3 capsid protein of adeno-associated  
PT virus, comprise a peptide containing a nucleus-shifting signal  
PT connected to its N-terminal -  
XX Claim 1; Page 14-16; 33p; Japanese.

XX The present sequence is provided in a specification relating to a  
CC virus-like particle-forming protein, and to a peptide containing a  
CC nuclear-shifting signal at its N-terminus that can form virus-like  
CC particles by shifting to the nucleus of the animal cell in which it  
CC is expressed. The method is used for forming virus-like particles  
CC from the VP3 capsid protein of adeno-associated virus (AAV).

XX Sequence 598 AA;

Query Match 84.9%; Score 2759.5; DB 22; Length 598;  
Best Local Similarity 82.8%; Pred. No. 8.4e-215;  
Matches 495; Conservative 43; Mismatches 59; Indels 1; Gaps 1;

QY 2 APGKKRPEVQSPQEPDSSSGIGKTGQOPAKKRLNFGQTDSESVDPDPLGEPATPAV 61  
DB 2 APGKKRPEVHSPVEPDSSSGTGAGQOPARKRLNFGQTDADSVDPDPLGPPAPPSGL 61  
QY 62 GPTTMASGGAPMADNNEGADGVGNASGNWCHDSTWLGDVITTTSTRTVALPTYNHLYK 121  
DB 62 GTNTMATGSGAPMADNNEGADGVGNSSGNWCHDSTWMDRVITTTSTRTVALPTYNHLYK 121  
QY 122 QISSASTGASNDNHFGYSTPWGYFDNFRFHCHFSPRDWORLNNNWGFRPKRLNFKLFN 181  
DB 122 QISSQS-GASNDNHFGYSTPWGYFDNFRFHCHFSPRDWORLNNNWGFRPKRLNFKLFN 180  
QY 182 IQVKEVTNDGVTIANNLSTVQVFSDESYQLPYVLGSAHQGLPPFPADVFMIPQYGY 241  
DB 181 IQVKEVTQNDGTTIANNLSTVQVFTDSEYQLPYVLGSAHQGLPPFPADVFMVPOYGY 240  
QY 242 LTLNNGSQAVGRSSFYCLEYFSPQMLRTGNNFTFSYTFEEVPHSSYAHQSOLDRLMNP 301  
DB 241 LTLNNGSQAVGRSSFYCLEYFSPQMLRTGNNFTFSYTFEDVFFHSSYAHQSOLDRLMNP 300  
QY 302 IDQYLYLNRTONQSGSAQNKDLLFSRGSFAGMSVQPKNMLPGPCYRQORVSKTKTDNN 361  
DB 301 IDQYLYLNRTONQSGTTTQSRLOFSQAGASDIRDQSRNMLPGPCYRQORVSKTSADNN 360  
QY 362 SNFTWTGASKYNLNGRESINPGTAMASHKDEDEKFFPMGVMIFGKESAGASNTALDNV 421  
DB 361 SEYSWTGATKYHLNGRDSLVPNPAMASHKDEKFFPQSGVLIFFKQOSEKTNVDIEKV 420  
QY 422 MITDEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMGALPGMWQDRDVLQGP 481  
DB 421 MITDEEIRTTNPVATEQYGSVSTNLQGRNQAATADVNTQGVLPGMWQDRDVLQGP 480  
QY 482 WAKIPHTDGHFHPSPMLMGFGLKNPPQILIKNTPVPANPFAFSATKFAFITQYSTGQ 541  
DB 481 WAKIPHTDGHFHPSPMLMGFGLKHPPQILIKNTPVPANPSTTFSAAKFAFITQYSTGQ 540

Qy	542	VSVEIEMWLOKENS	KRMNPEVOYTSN	AKSASVNDFTVDN	NGLYTEPRPIGR	RYLTRPL	599
				:		:	
Db	541	VSVEIEMWLOKENS	KRMNPEIOYTSN	AKSASVNDFTVD	NGVYSEPRPIGR	LYLTRNL	598

RESULT 12  
AAB50326

ID AAB50326 standard; protein; 734 AA.

AC AAB50326;

DT 09-MAR-2001 (first entry)

DE Adeno-associated virus capsid protein sequence.

KM Adeno-associated virus; AAV; capsid; virus binding inhibition; competitive inhibitor.

OS    Mastadenovirus.

OS Synthetic.

PN WO200073316-A2.

PD 07-DEC-2000.

PF 26-MAY-2000; 2000WO-US14466.

PR 28-MAY-1999; 99US-0321589.

PA (CELL-) CELL GENESYS INC.

PI Patel S, McArthur J;

DR WPI; 2001-061507/07.

PT New polypeptide portion of a virus protein binding to an antibody  
PT specific for the virus useful for inhibiting binding of the virus to a  
PT cell or for binding host antibody to provide a transient tolerant or  
PT non-responsive state -

PS Disclosure; Fig 2; 33pp; English.

The present sequence is given in a specification relating to polypeptide portions of a virus protein or its derivative, that bind to an antibody specific for the virus or inhibit binding of the virus to a cell. The polypeptides are used in inhibiting the binding of viruses to cells of a host. Oligopeptides that inhibit binding of virus to receptor can be used as competitive inhibitors to release bound virus in an adsorption-type assay, and if an antibody was used as an immunoadsorbent, the oligopeptide could be used to elute bound virus from a solid support to which virus antibody is immobilised. These oligopeptides may further be used to bind to host antibody to provide a transient tolerant or non-responsive state.

SQ Sequence 734 AA;

Query Match	84.2%	Score 2738;	DB 22;	Length 734;
Best Local Similarity	82.5%;	Pred. No. 6.2e-213;		
Matches 494;	Conservative 43;	Mismatches 60;	Indels 2;	Gaps 2;

Qy	1	TAPGKKRPVEQSPQEPDSSSGIGKTGQOPAKKRINFQOTGDSESVDPDQPLGEPAPTAA	60
			:
Db	138	TAPGKKRPVEHSPVEPDDSSSGTGAKGQOPAKKRINFQOTGDADSVDPDQPLGEPAPPSG	197
Qy	61	VGPPTMASGGGAPMADNNEGADGVGNASGNHCDSTWLGDVRVITSTRITWALPTYNNHLY	120
		:     :	:
Db	198	LGTNIMATGSGAPMADNNEGADGVGNSSGNHCDSTWNGDVRITSTRITWALPTYNNHLY	257
Qy	121	KQISSASTGASNDNHYFGYSTPWCYFDENRFHCHFSPRDWQRLINNNGWGRPKRLNFKLF	180
			:
Db	258	KQISSQS-GASNDNHYFGYSTPWCYFDENRFHCHFSPRDWQRLINNNGWGRPKRLNFKLF	316
Qy	181	NIQKEVTINDGVTTIANNLSTVQVFSDESYQLPYVLGSAHQGLPPEPADAVEMIPQYG	240

Db	317	NIQKEVTQNDGTTIANNLTSYQVETDSEYQLPYVLGSAHQGCLPFPADYFVWPQYG	376
QY	241	YLTLNNGSQAVGRSSFYCLEYFPDSQMLRTGNNTFTSYTFEEVPFHSSTAHQSQSLDRLMP	300
Db	377	YLTLNNGSQAVGRSSFYCLEYFPDSQMLRTGNNTFTSYTFEDVPFHSSTAHQSQSLDRLMP	436

```
QY      301 LIDÖLYLNRTONÖSGSAQNKDILFSGSPAGMSVÖPKNMLPGPCYRÖÖRVSKTKTDNN 360
        |||||:::||:|.:::|:|||||:|||||:|||||
DB      437 LIDÖLYLSRTNTPSGTTTQSRLOFSÖAGASDIRDÖSRNMLPGPCYRÖÖRVSKTSADNN 496
```

Qy 361 NSNFTWTGASKYNNLNGRESITNPGTAMASHKDEDEKFFPMSGVMI FGKSAGASNTALIDN 420  
| : : | : : | : : | : : | : : | : : | : : | : : | : : | : :  
Db 497 NSEYSWVGATKYHLNGRDSLVPDPAMASHKDEEKFPPQSGVLIFGKQSEKTNDIEK 556

```
Qy      421 VMITDEEEIKATNPVATERFGTVAVNFOSSSTDPATGVDHAMGALPGMWODRDVYLQGP   480
          |||||:::||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db     557 VMITDEEIRFTN-VATEQYGSYSTNLQRGNRQAATAADVNTQGVLPGMWMQDRDLYLQGP   615
```

```
QY      481 IWAKI PHTDGHFHPSPLMGGFGIKNPPQILIKNTVPANPAPAEFSATKASFITOYSTG 540
        ||||| :|||||
Db      616 IWAKI PHTDGHFHPSPLMGGFGIKHPPQILIKNTVPANPSTTFSAAKRASFITOYSTG 675
```

QY	541	ÖVSVEIEWELOKENS	KRWNP	EVQYTSN	AKSAND	FTVDN	NGLYTE	PRPIG	TRYLTRPL	599
			:			:				
Db	676	ÖVSVEIEWELOKENS <th>KRWNP</th> <th>EVQYTSN</th> <th>AKSAND</th> <th>FTVDN</th> <th>NGLYTE</th> <th>PRPIG</th> <th>TRYLTRNL</th> <th>734</th>	KRWNP	EVQYTSN	AKSAND	FTVDN	NGLYTE	PRPIG	TRYLTRNL	734

RESULT 13  
AAG65794

ID AAG65794 standard; Protein; 533 AA.

AC AAG65794;

DT 11-FEB-2002 (first entry)

DE Adeno-associated virus 2 (AAV-2) major coat protein VP3.

KM Recombinant viral vector; RVV; capsid; parvovirus; transgene; cytosstatic;  
KM inverted terminal repeat; nootropic; neuroprotective; antianemic; ITR;  
KM antidiabetic; antitumour; gene therapy; adeno-associated virus; AAV;  
KM major coat protein; AAV-2; VP3.

OS Adeno-associated virus 2.

PN W0200168888-A2.

PD 20-SEP-2001.

PF 13-MAR-2001; 2001WO-US079227.

PR 14-MAR-2000; 2000US-189110P.

PA (NEUR-) NEUROLOGIX INC.

PI Xiao W, During MJ;

DR WPI; 2001-596912/67.

DR N-PSDB; AAI66974.

PT Recombinant viral vector useful in improving gene therapy in a subject,  
PT and for increasing efficiency of entry into a cell, comprises a  
PT chimeric capsid having one non-native amino acid sequence and a desired  
PT transgene -

PS Disclosure; Page 51; 53pp; English.

CC The invention provides a recombinant viral vector (RVV) comprising a  
CC chimeric capsid (I) having at least one non-native amino acid sequence,  
CC derived from a capsid protein domain of parvovirus (II), a virus (III),  
CC or their combination, and a transgene flanked 5' and 3' by inverted  
CC terminal repeat (ITR) sequences, derived from (II), (III), or their  
CC combination. The RVV is useful for improving gene therapy in a subject

CC with a disorder, and for increasing the efficiency of entry into a cell,  
CC which involves producing (I) encapsulating a viral vector, and contacting  
CC a cell with RVV having (I) such that (I) binds to an attachment site on  
CC the cell surface and permits the vector to enter the cell efficiently.  
CC A pharmaceutical composition comprising RVV with (I) containing a  
CC transgene sequence associated with a disease or a disorder such that  
CC expression of the transgene would result in amelioration of the disease  
CC or disorder such as inherited neurological and metabolic diseases e.g.  
CC lysosomal storage disease, Lesch-Nyhan syndrome, amyloid neuropathy,  
CC Alzheimer amyloid, Duchenne's muscular dystrophy, diseases of the blood  
CC e.g. sickle-cell anemia, clotting disorders and thalassemias, cystic  
CC fibrosis, diabetes, diseases associated with hormone deficiencies,  
CC retinoblastoma and various types of neoplastic cells which include  
CC tumours especially central nervous system tumours, neoplasms, carcinomas,  
CC sarcomas, leukemias and lymphoma. The present sequence represents the  
CC adeno-associated virus 2 (AAV-2) major coat protein VP3. AAV-2 sequences  
CC are used in the construction of a chimeric vector.

XX Sequence 533 AA;

Query Match 76.3%; Score 2481.5; DB 22; Length 533;

Best Local Similarity 83.3%; Pred. No. 2.4e-192;

Matches 445; Conservative 37; Mismatches 51; Indels 1; Gaps 1;

OY 66 MASGGAPMADNNEGADGVGNASGNWHCDSTWLGDRVITSTRTWALPTYNHLYKQISS 125  
DB 1 MATSGAPMADNNEGADGVGNSSGNWHCDSTWNGDRVITSTRTWALPTYNHLYKQISS 60

OY 126 ASTGASNDNHFGYSTPWGYFDENRHHCHFSRDMQRLINNNMGFRPKRLNFKLFNIQVK 185  
DB 61 QS-GASNDNHFGYSTPWGYFDENRHHCHFSRDMQRLINNNMGFRPKRLNFKLFNIQVK 119

OY 186 EVTTNDGVTTIANNLTSTVQVFSDEYQLPYVLSAHQGCLEPPPADVFMIPQYGYLTN 245  
DB 120 EVTQNDGTTIANNLSTVQVFTDSEYQLPYVLSAHQGCLEPPPADVFMIPQYGYLTN 179

OY 246 NSQAVGRSSFYCLEYFPPSQMLRTGNNFTFSYTFEEVPHSSYAHQSQSLDRLMPLIDQY 305  
DB 180 NSQAVGRSSFYCLEYFPPSQMLRTGNNFTFSYTFEEDVPFHSSYAHQSQSLDRLMPLIDQY 239

OY 306 LYYLNRTONQSGSAQNKDLLFSRGSFAGMSVQPKNWLPGPCYRQQRVSKTKTDNNNSFT 365  
DB 240 LYYLSRTNTPSGTTQSRLOFSQAGASDIRQSRNWLPGPCYRQQRVSKTSADNNNSEYS 299

OY 366 WTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMGVMIFGKESAGASNTALDNVMTD 425  
DB 300 WTGATKYHLNGRDSLVPNPGPAMASHKDEDEKFFPQSGVLIFGKGSEKTNVDIEKMITD 359

OY 426 EBEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMGALPGMWQDRDYYLQGPIMAKI 485  
DB 360 EBEIRTNPVATEQYGSVSTNLQGRNRQAATADVNTQGVLPGMWQDRDYYLQGPIMAKI 419

OY 486 PHTDGHFHPSPMLMGFGGLKPNPQILIKNTPVPANPPAEFSATKFAFITQYSTGVSV 545  
DB 420 PHTDGHFHPSPMLMGFGGLKHPPOILIKNTPVPANPSTTFSIAKFAFITQYSTGVSV 479

OY 546 IEWELQKENSKRNPPEVQYTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 599  
DB 480 IEWELQKENSKRNPPEIQYTSNYKNSVNVDFTVDTNGVYSEPRPIGTRYLTRNL 533

RESULT 14

AAM51510 standard; protein; 533 AA.

XX AAM51510;

DT 02-JAN-2002 (first entry)

DE Adeno-associated virus VP3 capsid protein.

KW Adeno-associated virus; AAV; capsid; virus-like particle;

KW nuclear localisation signal; VP3.

XX Adeno associated virus.

XX JP2001169777-A.

XX 26-JUN-2001.

XX 30-JUL-1999; 99JP-0249140.

XX 30-JUL-1999; 99JP-0249140.

XX (HAND/) HANDA H.

XX WPI; 2001-599854/68.

PT New virus-like particles from VP3 capsid protein of adeno-associated  
PT virus, comprise a peptide containing a nucleus-shifting signal  
PT connected to its N-terminal -

PS Disclosure; Page 17-19; 33pp; Japanese.

CC The present sequence is provided in a specification relating to a  
CC virus-like particle-forming protein, and to a peptide containing a  
CC nuclear-shifting signal at its N-terminus that can form virus-like  
CC particles by shifting to the nucleus of the animal cell in which it  
CC is expressed. The method is used for forming virus-like particles  
CC from the VP3 capsid protein of adeno-associated virus (AAV).

SO Sequence 533 AA;

Query Match 76.2%; Score 2477.5; DB 22; Length 533;

Best Local Similarity 83.3%; Pred. No. 5e-192;

Matches 445; Conservative 36; Mismatches 52; Indels 1; Gaps 1;

OY 66 MASGGAPMADNNEGADGVGNASGNWHCDSTWLGDRVITSTRTWALPTYNHLYKQISS 125  
DB 1 MATSGAPMADNNEGADGVGNSSGNWHCDSTWNGDRVITSTRTWALPTYNHLYKQISS 60

OY 126 ASTGASNDNHFGYSTPWGYFDENRHHCHFSRDMQRLINNNMGFRPKRLNFKLFNIQVK 185  
DB 61 QS-GASNDNHFGYSTPWGYFDENRHHCHFSRDMQRLINNNMGFRPKRLNFKLFNIQVK 119

OY 186 EVTTNDGVTTIANNLTSTVQVFSDEYQLPYVLSAHQGCLEPPPADVFMIPQYGYLTN 245  
DB 120 EVTQNDGTTIANNLSTVQVFTDSEYQLPYVLSAHQGCLEPPPADVFMIPQYGYLTN 179

OY 246 NSQAVGRSSFYCLEYFPPSQMLRTGNNFTFSYTFEEVPHSSYAHQSQSLDRLMPLIDQY 305  
DB 180 NSQAVGRSSFYCLEYFPPSQMLRTGNNFTFSYTFEEDVPFHSSYAHQSQSLDRLMPLIDQY 239

OY 306 LYYLNRTONQSGSAQNKDLLFSRGSFAGMSVQPKNWLPGPCYRQQRVSKTKTDNNNSFT 365  
DB 240 LYYLSRTNTPSGTTQSRLOFSQAGASDIRQSRNWLPGPCYRQQRVSKTSADNNNSEYS 299

OY 426 EBEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMGALPGMWQDRDYYLQGPIMAKI 485  
DB 360 EBEIRTNPVATEQYGSVSTNLQGRNRQAATADVNTQGVLPGMWQDRDYYLQGPIMAKI 419

OY 486 PHTDGHFHPSPMLMGFGGLKPNPQILIKNTPVPANPPAEFSATKFAFITQYSTGVSV 545  
DB 420 PHTDGHFHPSPMLMGFGGLKHPPOILIKNTPVPANPSTTFSIAKFAFITQYSTGVSV 479

OY 546 IEWELQKENSKRNPPEVQYTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 599  
DB 480 IEWELQKENSKRNPPEIQYTSNYKNSVNVDFTVDTNGVYSEPRPIGTRYLTRNL 533

RESULT 15

AAM46313







GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 2003, 18:29:05 ; Search time 14.7426 Seconds  
(without alignments)  
1195.466 Million cell updates/sec

Title: US-09-807-802A-15  
Perfect score: 3251  
Sequence: 1 TAPGKKRPVEQSPQEPDSSS.....NNGLYTEPRPIGTRYLTRPL 599

Scoring table: BLOSUM62  
Gapop,10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/iaa/backfile1.pep:\*

Pred. NO. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Query length	DB ID	Description
1	1830.5	56.3	598	4	US-09-532-594B-16 Sequence 16, Appl
2	1830.5	56.3	734	4	US-09-532-594B-4 Sequence 4, Appl
3	1690.5	52.0	544	4	US-09-532-594B-18 Sequence 18, Appl
4	479.5	14.7	543	4	US-08-856-841-22 Sequence 22, Appl
5	326	10.0	415	4	US-08-856-841-20 Sequence 20, Appl
6	325	10.0	486	4	US-08-856-841-19 Sequence 19, Appl
7	325	10.0	500	4	US-08-856-841-16 Sequence 16, Appl
8	325	10.0	501	4	US-08-856-841-18 Sequence 18, Appl
9	318	9.8	395	4	US-08-856-841-13 Sequence 13, Appl
10	310	9.5	264	4	US-08-856-841-14 Sequence 14, Appl
11	303	9.3	398	4	US-08-856-841-21 Sequence 21, Appl
12	216.5	6.7	579	6	5223424-13 Patent No. 5223424
13	192	5.9	584	4	US-09-022-949-2 Sequence 2, Appl
14	178.5	5.5	387	4	US-08-856-841-17 Sequence 17, Appl
15	113.5	3.5	655	1	US-08-469-202-27 Sequence 27, Appl
16	113.5	3.5	655	2	US-08-469-202-27 Sequence 27, Appl
17	112.5	3.5	655	2	US-08-484-434C-34 Sequence 34, Appl
18	110.5	3.4	3060	2	US-08-487-826B-14 Sequence 14, Appl
19	110.5	3.4	624	3	US-08-947-965-78 Sequence 78, Appl
20	110.5	3.4	655	1	US-08-469-202-28 Sequence 28, Appl
21	109	3.4	655	2	US-08-484-434C-35 Sequence 35, Appl
22	109	3.4	1186	1	US-08-485-568A-4 Sequence 4, Appl
23	109	3.4	1186	1	US-08-357-698-6 Sequence 6, Appl
24	109	3.4	1186	2	US-08-590-554A-4 Sequence 4, Appl
25	109	3.4	1186	2	US-09-184-223-4 Sequence 4, Appl
26	106.5	3.3	1186	5	PCT-US93-12682-6 Sequence 6, Appl
27	106.5	3.3	1848	4	US-08-296-791-6 Sequence 6, Appl
			1848	5	PCT-US95-10661A-6 Sequence 6, Appl

28	104.5	3.2	717	4	US-09-626-589-1	Sequence 1, Appl
29	104	3.2	1013	4	US-09-415-522-8	Sequence 8, Appl
30	102	3.1	10182	4	US-09-134-001C-3159	Sequence 3159, Ap
31	101	3.1	1651	4	US-09-540-245A-18	Sequence 18, Appl
32	100	3.1	847	4	US-09-373-157-4	Sequence 4, Appl
33	99	3.0	824	4	US-09-626-589-3	Sequence 3, Appl
34	99	3.0	997	1	US-08-232-540-1	Sequence 1, Appl
35	99	3.0	997	1	US-08-428-949A-1	Sequence 1, Appl
36	99	3.0	997	1	US-08-428-948A-1	Sequence 1, Appl
37	99	3.0	997	2	US-08-428-946-1	Sequence 1, Appl
38	99	3.0	997	5	PCT-US95-04656-1	Sequence 1, Appl
39	99	3.0	998	1	US-08-233-008A-6	Sequence 6, Appl
40	99	3.0	1021	1	US-08-233-008A-2	Sequence 2, Appl
41	98.5	3.0	644	1	US-08-206-176-2	Sequence 2, Appl
42	98.5	3.0	1000	4	US-09-193-562D-30	Sequence 30, Appl
43	98.5	3.0	1036	4	US-08-891-640-3	Sequence 3, Appl
44	98.5	3.0	1091	3	US-08-633-768A-2	Sequence 2, Appl
45	98	3.0	1096	4	US-09-415-946-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1  
US-09-532-594B-16  
; Sequence 16, Application US/09532594B  
; Patent No. 6468524  
; GENERAL INFORMATION:  
; APPLICANT: Chorini, John A.  
; APPLICANT: Kocin, Robert M.  
; APPLICANT: Safier, Brian  
; APPLICANT: Davidson, Beverly  
; TITLE OF INVENTION: AAV4 VECTOR AND USES THEREOF  
; FILE REFERENCE: 14014.0252U2  
; CURRENT APPLICATION NUMBER: US/09/532,594B  
; CURRENT FILING DATE: 2000-03-22  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO 16  
; LENGTH: 598  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence; No. 6468524e =  
; NAME/KEY: misc feature  
; OTHER INFORMATION: AAV4 capsid protein VP2  
US-09-532-594B-16

Query Match	56.3%;	Score 1830.5;	DB 4;	Length 598;
Best Local Similarity	57.8%;	Pred. No. 2.6e-154;		
Matches	355;	Conservative 74;	Mismatches 154;	Indels 31; Gaps 10;
QY	1	TAPGKKRPVEQSPQEPDSSSGIGTKGQOPAKRLNF-GQTGDSSEVPDPPLGEPATPA	59	
DB	1	TAPGKKRPLIESPQOPDSSGTIGKKKQOPAKKLVFEDETGAAGDPPEGSTSG-----A	54	
QY	60	AVGPTTASGGGAPMADNNEGADGVGNASGNMHCSTWLGDRVITTTSTRTWALPTYNHL	119	
DB	55	MSDDSEMRAAAGAAVEGGQGDGVGNASGDWCHDSTWSEGHVTTTSTRTWALPTYNHL	114	
QY	120	YKOISSASTGASNDNHFGYFTWGYFDENRFHCHFSPRDWQRLINNMGFRPKRLNFKL	179	
DB	115	YKRIGE---SLQSNNTYNGFTWGYFDENRFHCHFSPRDWQRLINNMGMKPKAMRVKI	170	
QY	180	FNIOQKEVTNDGVTTIANNLSTVQVFSDEYQLPYVLGSAHQGLPPFPADVFMIPQY	239	
DB	171	FNIOQKEVTTSNGETVANNLSTVQIFADSSYELPYVMDAGQEGSLPPFPNDVFMVPOY	230	
QY	240	GY---LTLNNGSQAVGRSSFYCLEYFPSSQMLRTGNFTTSYTFEEVFPFSSYAHSGSLDR	296	
DB	231	GYCGLVTGNTSQOQTDNRNAPCYCLEYFPSSQMLRTGNFTTSYTFEEVFPFSSYAHSGSLDR	290	



[illegible]

## RESULT 2

```

US-09-532-594B-4
; Sequence 4, Application US/09532594B
; Patent No. 6468524.
; GENERAL INFORMATION:
; APPLICANT: Chorini, John A.
; APPLICANT: Kotin, Robert M.
; APPLICANT: Safer, Brian
; APPLICANT: Davidson, Beverly
; TITLE OF INVENTION: 'AAV4 VECTOR AND USES THEREOF
; FILE REFERENCE: 14014.025202
; CURRENT APPLICATION NUMBER: US/09/532,594B
; CURRENT FILING DATE: 2000-03-22
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 734
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence; No. 6468524e =
; NAME/KEY: misc feature
; OTHER INFORMATION: AAV4 capsid protein VP1
US-09-532-594B-4

```

```

Query Match      56.3%; Score 1830.5; DB 4; Length 734;
Best Local Similarity 57.8%; Pred. No. 3.6e-154;
Matches 355; Conservative 74; Mismatches 154; Indels 31; Gaps 10;

QY      1  TAPGKKRPVEQSPQEPDSSGIGIKTGQQPAKRLNF-GQGTGHSSEVPDPQPLGEPATPA 59
      |||||:::|||||:|||||:|||||:|::|:|:|
Dd      137 TAPGKKRPLIESPQQPDSSTGIGKKGQPAKKLVFEDEGTGADGCPREGSTSG-----A 190
QY      60  AVGPTTMAAGGAPMADNNEGADGVGNASGNWHCDSTWLGDFVITTSRTWALPTYNNHL 119
      :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Dd      191 MSDSEMRAAAGAAVEGGQGDAGVGNASGDWHCDSTWSEGHVITTSRTWVLPPTYNNHL 250
QY      120 YKQISSASTGASNDNHFGYSTPWGYFDENRFCHFSPRDWQRLINNNWGFRLPKRLNPKL 179
      ||:::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Dd      251 YKRLGE----SLQSNITYNGFSTPWGYFDENRFCHFSPRDWQRLINNNWGMRLPKAMRVKI 306
QY      180 FNIQVKEVTINDGVTTIANNLSTVQVFSDEYQLPYVLGSLHQGLPPFPADVFMIPQY 239
      |||||:::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Dd      307 FNIQVKEVTTNGETIVANNLSTVQIFADSSYELPYVMDAQEGESLPPFPNDVFMVPOY 366
QY      240 GY---LTLNNGSQAVGRSSFYCLEYFPSQMLRTGNFTFSYTFEEVFPHSSYAHSQSLDR 296
      |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

```

Db	367	GYCGLVTGNTSQQOQTDNRNAYCLEYFPSPQMLRTGNNFETYSFEKVPFHSHYAHQSQSLDR	426
QY	297	LMNPLIDQYLYYLNRTQN---QSGSAQNKDLLFSRGSPPAGMSVQPKMWLPQPCYRQORV	352
Db	427	LMNPLIDQYLMGLQSTTGTITLNAGTATTN---FTKLRPPTNFNFKKQWLPQPSIKQGF	483
QY	353	SKTKTDNNNSNFTWTGAS---KY---NLNGRESIINPGTAMASHKDEDEKFFPMGVM	405
Db	484	SKTA--NONYKIPATGSDSLIKYETHSTLTDGRWSALTTPGPPMATAGPADSK-FSNSQLIF	540
QY	406	FGKESAGASNTALDNVMTDEEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMGAL	465
Db	541	AGPKONGNTATVPCTLIFTSEEEIATNATDTDMWGNLPGGDQSNSLPIYDRLTALGAV	600
QY	466	PGMWQDRDLYLQGPIMAKIPHTDGHFHPSPPLMGFGFLKNPPQILIKNTBPVANPPAEF	525
Db	601	PGMWQNRDIYYQGPIMAKIPHTDGHFHPSPPLIGFGFLKHPPIQIFIKNTBPVANPATTF	660
QY	526	SATKFAFITQYSTGQVSVEIEMELOKENSKRWNPEVQYTSNYAKSANVDFTVDNNGLYT	585
Db	661	SSTPWNSTFYQYSTGQVSQVIDMEIQEKERSKRWNPEVQFTSNYGGQNSLMLMAPDAAGKYT	720
QY	586	EPRPIGTRYLTRPL 599	
Db	721	EPRAIGTRYLTHHL 734	

### RESULT 3

```

US-09-532-594B-18
; Sequence 18, Application US/09532594B
; Patent No. 6468524
; GENERAL INFORMATION:
; APPLICANT: Chorini, John A.
; APPLICANT: Kotin, Robert M.
; APPLICANT: Safer, Brian
; APPLICANT: Davidson, Beverly
; TITLE OF INVENTION: AAV4 VECTOR AND USES THEREOF
; FILE REFERENCE: 14014.0252U2
; CURRENT APPLICATION NUMBER: US/09/532,594B
; CURRENT FILING DATE: 2000-03-22
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 544
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence; No. 6468524e =
; OTHER INFORMATION: synthetic construct
; NAME/KEY: misc feature
; OTHER INFORMATION: AAV4 capsid protein VP3
US-09-532-594B-18

```

[illegible]



```

RESULT 5
US-08-856-841-20
; Sequence 20, Application US/08856841
; Patent No. 6274307
; GENERAL INFORMATION:
; APPLICANT: ERWIN SOUTSCHEK
; APPLICANT: MANFRED MOTZ
; TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES
; TITLE OF INVENTION: OR POLYPEPTIDES FROM THE PARVOVIRUS B19
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROOKS HAIDT HAFNER & DELAUNTY
; STREET: 99 PARK AVENUE
; CITY: NEW YORK
; STATE: NY
; COUNTRY: USA
; ZIP: 10016
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" FLOPPY DISC
; COMPUTER: AT&T - IBM COMPATIBLE
; OPERATING SYSTEM: MS-DOS Version 6.2
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/856, 841
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/214, 658
; FILING DATE: 16-MARCH-1994
; APPLICATION NUMBER: US 07/917, 096
; FILING DATE: 4-AUGUST-1992
; APPLICATION NUMBER: PCT/DE91/00106
; FILING DATE: 8-FEBRUARY-1991
; APPLICATION NUMBER: DE40038262
; FILING DATE: 8-FEBRUARY-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: ROBINSON, WILLIAM R.
; REGISTRATION NUMBER: 27,224
; REFERENCE/DOCKET NUMBER: LKR-9222-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 697-3355
; TELEFAX: (212) 557-5635
; TELEX: NONE
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 415
; TYPE: AMINO ACID
; TOPOLOGY: LINEAR
; MOLECULE TYPE:
; DESCRIPTION: PEPTIDE
; HYPOTHETICAL: N/A
; ANTI-SENSE: N/A
; FRAGMENT TYPE: INTERNAL
; ORIGINAL SOURCE: SERUM FROM PATIENT WITH ACUTE
; ORIGINAL SOURCE: INFECTION (ERYTHEMA INFECTIOSUM)
; IMMEDIATE SOURCE: GENETICALLY ENGINEERED PEPTIDE
; POSITION IN GENOME: N/A
; FEATURE:
; NAME/KEY: N/A
; LOCATION: N/A
; IDENTIFICATION METHOD: amino acid analysis and
; IDENTIFICATION METHOD: mass spectrometry
; OTHER INFORMATION:
; PUBLICATION INFORMATION:
; AUTHORS: COSSART, Y.E.
; AUTHORS: FIELD, A.M.
; AUTHORS: CANT, B.
; AUTHORS: WIDDOWS, D.
; TITLE: PARVOVIRUS-LIKE PARTICLES IN HUMAN SERA
; JOURNAL: LANCET
; VOLUME: I
; ISSUE:
; PAGES: 72 - 73

```

```

; DATE: 1975
; DOCUMENT NUMBER:
; FILING DATE:
; PUBLICATION DATE:
; RELEVANT RESIDUES IN SEQ ID NO: 20:
US-08-856-841-20

Query Match      10.0%; Score 326; DB 4; Length 415;
Best Local Similarity 26.0%; Pred. No. 1e-20;
Matches 101; Conservative 66; Mismatches 180; Indels 42; Gaps 13;

QY 193 VTTIANLTSTVOVFSDSYOLPYVLGSAHQCLPPPADVMIPQYGLTLNN-GSQAV 251
   :|::||::||::||::||::||::||::||::||::||::||::||::||::||:
Db 1 MTMITPSLHACMLV--DHEKKYPVLGGQGDTLAPELPIWVFPPOYAYLTVGDVNTQGI 58

QY 252 G-----RSSFYCLEYFPSOMLRGTGNFTFSTFEVEVPFHSSYAHQSOLDRLMPLI 302
   :|::||::||::||::||::||::||::||::||::||::||::||::||:
Db 59 SGDSKSLASEESAFLVLEHSSFOLLGTGTASMSYKPPVPENLEGCSQHFEYMNPPL- 117

QY 303 DQYLYYLNRTQNQSGSQAQNKDLLFSRGSFAGMSVOPKNWLPGPCYRQQRVSKTKTDNNS 362
   |::||::||::||::||::||::||::||::||::||::||::||::||:
Db 118 --YGRLLGVPTLGDDPKFRSL-----THEDHALDPQNMFGPPLVNSVSTKEGDSSTGA 170

QY 363 NFWTGWASKYNLNGRESILNPG-TAMASHKDDEKFFPMMSGVMI FGKESAGASNTALDNV 421
   ||::||::||::||::||::||::||::||::||::||::||::||:
Db 171 GKALTGLSTGSONTRISLRPGPVSOQYHHWDIDKYVTGINAISHGQTTYG--NAEDKE 227

QY 422 -----MTDEEIKATNPVATERFGTVAVNFOSSSTDPAIGDVHAMGALPGMWODRD 474
   ::||::||::||::||::||::||::||::||::||::||::||:
Db 228 YQQGVGRFPNEKEQLKQLGLNMHTY-----FPNKGTOOYTDOIE-RPLMVGSVMNRRA 280

QY 475 VYLOGPIWAKIPHDTGHFHPSP-PLMGFGFKLPNPPOILLIKNTVPANPPAEFSATKFASF 533
   ::||::||::||::||::||::||::||::||::||::||::||:
Db 281 LHYESQLWSKIPLNDSEFKTFQFALGGWGHLHOPPQIFLK--ILPESGPIGIKSMGIT 338

QY 534 ITQYSTGVSVSEIEWEL-QEKNSKRWNPE 561
   :||::||::||::||::||::||::||::||::||::||::||:
Db 339 LVQYAVGIMTVMTFKLGPRKATGRWNPQ 367

RESULT 6
US-08-856-841-19
; Sequence 19, Application US/08856841
; Patent No. 6274307
; GENERAL INFORMATION:
; APPLICANT: ERWIN SOUTSCHEK
; APPLICANT: MANFRED MOTZ
; TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES
; TITLE OF INVENTION: OR POLYPEPTIDES FROM THE PARVOVIRUS B19
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROOKS HAIDT HAFNER & DELAHUNTY
; STREET: 99 PARK AVENUE
; CITY: NEW YORK
; STATE: NY
; COUNTRY: USA
; ZIP: 10016
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" FLOPPY DISC
; COMPUTER: AT&T - IBM COMPATIBLE
; OPERATING SYSTEM: MS-DOS Version 6.2
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/856, 841
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/214, 658
; FILING DATE: 16-MARCH-1994
; APPLICATION NUMBER: US 07/917, 096
; FILING DATE: 4-AUGUST-1992
; APPLICATION NUMBER: PCT/DE91/00106
; FILING DATE: 8-FEBRUARY-1991

```





DOCUMENT NUMBER:  
FILING DATE:  
PUBLICATION DATE:  
RELEVANT RESIDUES IN SEQ ID NO: 16:  
PUBLICATION INFORMATION:  
AUTHORS: MANIATIS, T.  
AUTHORS: FRITSCH, E.F.  
AUTHORS: SAMBROOK, J.  
TITLE: MOLECULAR CLONING  
JOURNAL: COLD SPRING HARBOR, NY  
VOLUME:  
ISSUE:  
PAGES:  
DATE: 1982  
DOCUMENT NUMBER:  
FILING DATE:  
PUBLICATION DATE:  
RELEVANT RESIDUES IN SEQ ID NO: 16:  
PUBLICATION INFORMATION:  
AUTHORS: SMITH, D.B.  
AUTHORS: JOHNSON, K.S.  
TITLE: SINGLE STEP PURIFICATION OF POLYPEPTIDES  
TITLE: EXPRESSED IN ESCHERICHIA COLI AS FUSIONS WITH  
TITLE: GLUTATHIONE S. TRANSFERASE  
JOURNAL: GENE  
VOLUME:  
ISSUE: 67  
PAGES: 31 - 40  
DATE: 1988  
DOCUMENT NUMBER:  
FILING DATE:  
PUBLICATION DATE:  
RELEVANT RESIDUES IN SEQ ID NO: 16:  
US-08-856-841-16

Query Match 10.0%; Score 325; DB 4; Length 500;  
Best Local Similarity 32.2%; Pred. No. 1.7e-20;  
Matches 88; Conservative 34; Mismatches 127; Indels 24; Gaps 6;

QY 51 LGPPATPAVGTPTMASGGAPMADNNEGADVGNASGNMHCSTWLGDRVITSTRTW 110  
DB 221 LPEVPAYNASEKYPSTSVNSAE-ASTGAGGGGNSVKSMSEGATFSANSTCTFSRQF 279  
QY 111 ALPTYNMHLKQISSASTGASND-----NHFGYSTPWGYPDNRFHCHFSRDM 160  
DB 280 LIPYDPEHHYKVPSPASSCHNASGKEAKVCTISPIMGYSTPWRYLDNALNLFSPLEF 339  
QY 161 QRLINNNGFRPKLNFKLNIOVKEVT--TNDGVTIANNLSTVQVFSDESEYQLPYVL 218  
DB 340 QHLIENYGSIAIPDALVTITSEIAVKDVTDKTGGV-QVTDSTGRCLMLVDHEKYPYVL 398  
QY 219 GSAHQCLPPFPADVFMIPQYGYLTINN-GSQAVG-----RSSPYCLEYFPSPQMLR 268  
DB 399 GQGODTIAPELPIMWYFPQYAYLTVGDVNTGSGDSKSLASEESAFYVLEHSSQQLG 458  
QY 269 TGNNTFSYTFEEVFPFHSSYAHQSGLDRMLNPL 301  
DB 459 TGGTASMSYKFPVPPEPNLEGCSDHFYEMYNPL 491

RESULT 8  
US-08-856-841-18  
Sequence 18, Application US/08856841  
Patent No. 6274307  
GENERAL INFORMATION:  
APPLICANT: ERWIN SOUTSCHEK  
APPLICANT: MANFRED MOTZ  
TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES  
TITLE OF INVENTION: OR POLYPEPTIDES FROM THE PARVOVIRUS B19  
NUMBER OF SEQUENCES: 28  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROOKS HAIDT HAFNER & DELAHUNTY  
STREET: 99 PARK AVENUE

CITY: NEW YORK  
STATE: NY  
COUNTRY: USA  
ZIP: 10016  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" FLOPPY DISC  
COMPUTER: AT&T - IBM COMPATIBLE  
OPERATING SYSTEM: MS-DOS Version 6.2  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/856,841  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/214,658  
FILING DATE: 16-MARCH-1994  
APPLICATION NUMBER: US 07/917,096  
FILING DATE: 4-AUGUST-1992  
APPLICATION NUMBER: PCT/DE91/00106  
FILING DATE: 8-FEBRUARY-1991  
APPLICATION NUMBER: DE40038262  
FILING DATE: 8-FEBRUARY-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: ROBINSON, WILLIAM R.  
REGISTRATION NUMBER: 27,224  
REFERENCE/DOCKET NUMBER: LKR-9222-A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 697-3355  
TELEFAX: (212) 557-5635  
TELEX: NONE  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 501  
TYPE: AMINO ACID  
TOPOLOGY: LINEAR  
MOLECULE TYPE:  
DESCRIPTION: PEPTIDE  
HYPOTHETICAL: N/A  
ANTI-SENSE: N/A  
FRAGMENT TYPE: INTERNAL  
ORIGINAL SOURCE: SERUM FROM PATIENT WITH ACUTE  
ORIGINAL SOURCE: INFECTION (ERYTHEMA INFECTIONOSUM)  
IMMEDIATE SOURCE: GENETICALLY ENGINEERED PEPTIDE  
POSITION IN GENOME: N/A  
FEATURE:  
NAME/KEY: N/A  
LOCATION: N/A  
IDENTIFICATION METHOD: amino acid analysis and  
IDENTIFICATION METHOD: mass spectrometry  
OTHER INFORMATION:  
PUBLICATION INFORMATION:  
AUTHORS: COSART, Y.E.  
AUTHORS: FIELD, A.M.  
AUTHORS: CANT, B.  
AUTHORS: WIDDOWS, D.  
TITLE: PARVOVIRUS-LIKE PARTICLES IN HUMAN SERA  
JOURNAL: LANCET  
VOLUME: 1  
ISSUE:  
PAGES: 72 - 73  
DATE: 1975  
DOCUMENT NUMBER:  
FILING DATE:  
PUBLICATION DATE:  
RELEVANT RESIDUES IN SEQ ID NO: 18:  
US-08-856-841-18

Query Match 10.0%; Score 325; DB 4; Length 501;  
Best Local Similarity 32.2%; Pred. No. 1.7e-20;  
Matches 88; Conservative 34; Mismatches 127; Indels 24; Gaps 6;

QY 51 LGPPATPAVGTPTMASGGAPMADNNEGADVGNASGNMHCSTWLGDRVITSTRTW 110

Db 222 LPEVPAYNASEKYPMSVNSAE-ASTGAGGGGNSVKSMWSEGATFSANSVTCTFSRQF 280  
QY 111 ALPTYNHLYKQISSASTGASND-----NHFGYSTPWGYEDENRFCHFSPRDW 160  
Db 281 LIPYDEHHYKVPSPASSCHNAGSKAKVCTISPIMGYSTPWRJLDENALNLFSPLEF 340  
QY 161 QRLINNNWGFRPKRLNFKLFNIQVKEVT--TNDGVTIANNLTSFVQVFSDESEYQLPYVL 218  
Db 341 QHLIENYGSIAKDALTVTISEIAVKDVTDTGGGV-QVTDSTGRLCMLVDHEKXPYVL 399  
QY 219 GSAHQGCLPPFPADVEMIPQYGLTLNN-GSQAVG-----RSSFYCLEYFPSPQMR 268  
Db 400 GQGQDTLAPELPIWVYFPQYAYLTVGDVNTQGISGDSKLAESASAFYVLEHSSFOLLG 459  
QY 269 TGNNTFSYTFEEVFPFHSSYAHQSGLDRLNPL 301  
Db 460 TGGTASMSYKFPVPPENLGCSQHFEYENPL 492

## RESULT 9

US-08-856-841-13  
; Sequence 13, Application US/08856841  
; Patent No. 6274307  
; GENERAL INFORMATION:  
; APPLICANT: ERWIN SOUTSCHEK  
; APPLICANT: MANFRED MOTZ  
; TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES  
; TITLE OF INVENTION: OR POLYPEPTIDES FROM THE PARVOVIRUS B19  
; NUMBER OF SEQUENCES: 28  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROOKS HAIDT HAFNER & DELAHUNTY  
; STREET: 99 PARK AVENUE  
; CITY: NEW YORK  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10016  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" FLOPPY DISC  
; COMPUTER: AT&T - IBM COMPATIBLE  
; OPERATING SYSTEM: MS-DOS Version 6.2  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/856,841  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/214,658  
; FILING DATE: 16-MARCH-1994  
; APPLICATION NUMBER: US 07/917,096  
; FILING DATE: 4-AUGUST-1992  
; APPLICATION NUMBER: PCT/DE91/00106  
; FILING DATE: 8-FEBRUARY-1991  
; APPLICATION NUMBER: DE40038262  
; FILING DATE: 8-FEBRUARY-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: ROBINSON, WILLIAM R.  
; REGISTRATION NUMBER: 27,224  
; REFERENCE/DOCKET NUMBER: LKR-9222-A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 697-3355  
; TELEFAX: (212) 557-5635  
; TELEX: NONE  
; INFORMATION FOR SEQ ID NO: 13:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 395  
; TYPE: AMINO ACID  
; TOPOLOGY: LINEAR  
; MOLECULE TYPE: PEPTIDE  
; DESCRIPTION: N/A  
; HYPOTHETICAL: N/A  
; ANTI-SENSE: N/A  
; FRAGMENT TYPE: INTERNAL  
; ORIGINAL SOURCE: SERUM FROM PATIENT WITH ACUTE

; ORIGINAL SOURCE: INFECTION (ERYTHEMA INFECTIOSUM)  
; IMMEDIATE SOURCE: GENETICALLY ENGINEERED PEPTIDE  
; POSITION IN GENOME: N/A  
; FEATURE:  
; NAME/KEY: N/A  
; LOCATION: N/A  
; IDENTIFICATION METHOD: amino acid analysis and  
; IDENTIFICATION METHOD: mass spectrometry  
; OTHER INFORMATION:  
; PUBLICATION INFORMATION:  
; AUTHORS: COSSART, Y.E.  
; AUTHORS: FIELD, A.M.  
; AUTHORS: CANT, B.  
; AUTHORS: WIDOWS, D.  
; TITLE: PARVOVIRUS-LIKE PARTICLES IN HUMAN SERA  
; JOURNAL: LANCET  
; VOLUME: 1  
; ISSUE:  
; PAGES: 72 - 73  
; DATE: 1975  
; DOCUMENT NUMBER:  
; FILING DATE:  
; PUBLICATION DATE:  
; RELEVANT RESIDUES IN SEQ ID NO: 13:  
; PUBLICATION INFORMATION:  
; AUTHORS: MANIATIS, T.  
; AUTHORS: FRITSCH, E.F.  
; AUTHORS: SAMERCOCK, J.  
; TITLE: MOLECULAR CLONING  
; JOURNAL: COLD SPRING HARBOR, NY  
; VOLUME:  
; ISSUE:  
; PAGES:  
; DATE: 1982  
; DOCUMENT NUMBER:  
; FILING DATE:  
; PUBLICATION DATE:  
; RELEVANT RESIDUES IN SEQ ID NO: 13:  
; PUBLICATION INFORMATION:  
; AUTHORS: SMITH, D.B.  
; AUTHORS: JOHNSON, K.S.  
; TITLE: SINGLE STEP PURIFICATION OF POLYPEPTIDES  
; TITLE: EXPRESSED IN ESCHERICHIA COLI AS FUSIONS WITH  
; TITLE: GLUTATHIONE S. TRANSFERASE  
; JOURNAL: GENE  
; VOLUME:  
; ISSUE: 67  
; PAGES: 31 - 40  
; DATE: 1988  
; DOCUMENT NUMBER:  
; FILING DATE:  
; PUBLICATION DATE:  
; RELEVANT RESIDUES IN SEQ ID NO: 13:  
; US-08-856-841-13

Query Match 9.8%; Score 318; DB 4; Length 395;  
Best Local Similarity 25.4%; Pred. No. 4.8e-20;  
Matches 99; Conservative 61; Mismatches 167; Indels 62; Gaps 13;

QY 193 VTTIANNLSTVQVFSDEYQLPYVLSAHQGCLPPFPADVEMIPQYGLTLNN-GSQAV 251  
Db 1 MTMITPSLAACMLV--DHEYKYPYVLGQGDTLAPELPIWVYFPQYAYLTVGDVNTQGI 58  
QY 252 G-----RSSFYCLEYFPSPQMRGTGNNFTFSYTFEEVFPFHSSYAHQSGLDRLNPLI 302  
Db 59 SGDSKLAESASAFYVLEHSSFOLLGTGTASMSYKFPVPPENLGCSQHFEYENPL- 117  
QY 303 DQYLYYLNRTQNSGAQNKDLLFSRGSFAGMSVQPKWLLPGPCYRQQRVSKTKTDNNNS 362  
Db 118 --YGSRLGVPDTLGGDPKFRSL-----THEDHAIQPNFMFPLVNSVSTKEGSSNTGA 170  
QY 363 NPTWTGASKYNLNGRESITNDG-TAMASHKDEDEKFPFMSGVMIFGKESAGASNTALDNV 421



Db 171 GKALTGLSTGTSQNTFRSLRPGVVSQPYHHMDPKVVTGINAISHGQTFY---NAEDKE 227  
QY 422 -----MITDEEIKATNPVATERFGTVAVNFQSSSTDPATGVDVHANGALPGMWQDRD 474  
Db 228 YQGGVGRFPNEKEQLKQLOGLMHTY-----FPNKGTOQYTDQIE-RPLMWGSVMNRR 280  
QY 475 VYLQGPIMAKIPHTDGFHPS-PLMGFGGLKNPPQILIKNTVPANPPAEFSATKFAF 533  
Db 281 LHYESQLMSKIPNLDSTFKTQFALGCGWGLHQPPOI-----F 318  
QY 534 ITQYSTQVSVIEIWEI-OKENSKRWPE 561  
Db 319 LKQYAVGIMTVMTFRKLGPBKATGRWNPQ 347

RESULT 10  
US-08-856-841-14  
Sequence 14, Application US/08856841  
Patent No. 6274307  
GENERAL INFORMATION:  
APPLICANT: ERWIN SOUTSCHEK  
APPLICANT: MANFRED MOTZ  
TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES  
TITLE OF INVENTION: OR POLYPEPTIDES FROM THE PARVOVIRUS B19  
NUMBER OF SEQUENCES: 28  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROOKS HAIDT HAFNER & DELAHUNTY  
STREET: 99 PARK AVENUE  
CITY: NEW YORK  
STATE: NY  
COUNTRY: USA  
ZIP: 10016  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" FLOPPY DISC  
COMPUTER: AT&T - IBM COMPATIBLE  
OPERATING SYSTEM: MS-DOS Version 6.2  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/856,841  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/214,658  
FILING DATE: 16-MARCH-1994  
APPLICATION NUMBER: US 07/917,096  
FILING DATE: 4-AUGUST-1992  
APPLICATION NUMBER: PCT/DE91/00106  
FILING DATE: 8-FEBRUARY-1991  
APPLICATION NUMBER: DE40038262  
FILING DATE: 8-FEBRUARY-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: ROBINSON, WILLIAM R.  
REGISTRATION NUMBER: 27,224  
REFERENCE/DOCKET NUMBER: LKR-9222-A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 697-3355  
TELEFAX: (212) 557-5635  
TELEX: NONE  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 264  
TYPE: AMINO ACID  
TOPOLOGY: LINEAR  
MOLECULE TYPE:  
DESCRIPTION: PEPTIDE  
HYPOTHETICAL: N/A  
ANTI-SENSE: N/A  
FRAGMENT TYPE: INTERNAL  
ORIGINAL SOURCE: SERUM FROM PATIENT WITH ACUTE  
ORIGINAL SOURCE: INFECTION (ERYTHEMA INFECTIOSUM)  
IMMEDIATE SOURCE: GENETICALLY ENGINEERED PEPTIDE  
POSITION IN GENOME: N/A  
FEATURE:

NAME/KEY: N/A  
LOCATION: N/A  
IDENTIFICATION METHOD: amino acid analysis and  
IDENTIFICATION METHOD: mass spectrometry  
OTHER INFORMATION:  
PUBLICATION INFORMATION:  
AUTHORS: COSSART, Y.E.  
AUTHORS: FIELD, A.M.  
AUTHORS: CANT, B.  
AUTHORS: WIDOWS, D.  
TITLE: PARVOVIRUS-LIKE PARTICLES IN HUMAN SERA  
JOURNAL: LANCET  
VOLUME: I  
ISSUE:  
PAGES: 72 - 73  
DATE: 1975  
DOCUMENT NUMBER:  
FILING DATE:  
PUBLICATION DATE:  
RELEVANT RESIDUES IN SEQ ID NO: 14:  
PUBLICATION INFORMATION:  
AUTHORS: MANIATIS, T.  
AUTHORS: FRITSCH, E.F.  
AUTHORS: SAMBROOK, J.  
TITLE: MOLECULAR CLONING  
JOURNAL: COLD SPRING HARBOR, NY  
VOLUME:  
ISSUE:  
PAGES:  
DATE: 1982  
DOCUMENT NUMBER:  
FILING DATE:  
PUBLICATION DATE:  
RELEVANT RESIDUES IN SEQ ID NO: 14:  
PUBLICATION INFORMATION:  
AUTHORS: SMITH, D.B.  
AUTHORS: JOHNSON, K.S.  
TITLE: SINGLE STEP PURIFICATION OF POLYPEPTIDES  
TITLE: EXPRESSED IN ESCHERICHIA COLI AS FUSIONS WITH  
TITLE: GLUTATHIONE S. TRANSFERASE  
JOURNAL: GENE  
VOLUME:  
ISSUE: 67  
PAGES: 31 - 40  
DATE: 1988  
DOCUMENT NUMBER:  
FILING DATE:  
PUBLICATION DATE:  
RELEVANT RESIDUES IN SEQ ID NO: 14:  
US-08-856-841-14  
Query Match 9.5%; Score 310; DB 4; Length 264;  
Best Local Similarity 31.3%; Pred. No. 1.3e-19;  
Matches 78; Conservative 33; Mismatches 104; Indels 34; Gaps 6;  
QY 57 TPAVGPPTMASGGGAPMADNNEGADGVGNASGNMHCSTWLGDRVITSTRTWALPTVN 116  
Db 5 TPSLAAEASTGAGG-----GSNSVKSMWSEGFANSVTCTFSRQFLIPYDP 53  
QY 117 NMLYKQISSASTGASND-----NHYFGYSTPWGTFDFNRFFCHFSRDMQRLINN 166  
Db 54 EHHYKVSPPAASSCHNASGKEAKVCTISPIMGYSTPWRYLDENALNLFSPLEFQHLIEN 113  
QY 167 NWGFRPKRLNFKLFNIQVKEVT--TNDGVTTIANLSTVQVFSDSYQLPYVLSAHQG 224  
Db 114 YGSIAPDALVTITSEIAVKDVTDTKGGV-QVTDSTTGRCLMVDHEKYKPYVLGGQDGT 172  
QY 225 CLPFPADVFEMIPQGYLTLLN-GSQAVG-----RSGFYCLEYFSPSOMLRGTGNNFT 274  
Db 173 LAPELPIMVYFPPQYAYLTVGDVNTQGISGDSKSLASESAFYVLEHSSFOLLGTGTAS 232  
QY 275 FSYTFEEYVP 283







;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: BROOKS HAIDT HAFNER & DELAHUNTY  
;; STREET: 99 PARK AVENUE  
;; CITY: NEW YORK  
;; STATE: NY  
;; COUNTRY: USA  
;; ZIP: 10016  
;;  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: 3.5" FLOPPY DISC  
;; COMPUTER: AT&T - IBM COMPATIBLE  
;; OPERATING SYSTEM: MS-DOS Version 6.2  
;; SOFTWARE: ASCII  
;;  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/856,841  
;; FILING DATE:  
;; CLASSIFICATION:  
;;  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/214,658  
;; FILING DATE: 16-MARCH-1994  
;; APPLICATION NUMBER: US 07/917,096  
;; FILING DATE: 4-AUGUST-1992  
;; APPLICATION NUMBER: PCT/DE91/00106  
;; FILING DATE: 8-FEBRUARY-1991  
;; APPLICATION NUMBER: DE40038262  
;; FILING DATE: 8-FEBRUARY-1990  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: ROBINSON, WILLIAM R.  
;; REGISTRATION NUMBER: 27,224  
;; REFERENCE/DOCKET NUMBER: LKR-9222-A  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (212) 697-3355  
;; TELEFAX: (212) 557-5635  
;; TELEX: NONE  
;;  
;; INFORMATION FOR SEQ ID NO: 17:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 387  
;; TYPE: AMINO ACID  
;; TOPOLOGY: LINEAR  
;;  
;; MOLECULE TYPE:  
;; DESCRIPTION: PEPTIDE  
;; HYPOTHETICAL: N/A  
;; ANTI-SENSE: N/A  
;;  
;; FRAGMENT TYPE: INTERNAL  
;; ORIGINAL SOURCE: SERUM FROM PATIENT WITH ACUTE  
;; IMMEDIATE SOURCE: INFECTION (ERYTHEMA INFECTION)  
;; POSITION IN GENOME: N/A  
;;  
;; FEATURE:  
;; NAME/KEY: N/A  
;; LOCATION: N/A  
;; IDENTIFICATION METHOD: amino acid analysis and  
;; IDENTIFICATION METHOD: mass spectrometry  
;;  
;; OTHER INFORMATION:  
;; PUBLICATION INFORMATION:  
;; AUTHORS: COSSART, Y.E.  
;; AUTHORS: FIELD, A.M.  
;; AUTHORS: CANT, B.  
;; AUTHORS: WIDDOWS, D.  
;; TITLE: PARVOVIRUS-LIKE PARTICLES IN HUMAN SERA  
;; JOURNAL: LANCET  
;; VOLUME: 1  
;; ISSUE:  
;; PAGES: 72 - 73  
;; DATE: 1975  
;; DOCUMENT NUMBER:  
;; FILING DATE:  
;; PUBLICATION DATE:  
;; RELEVANT RESIDUES IN SEQ ID NO: 17:  
;; PUBLICATION INFORMATION:  
;; AUTHORS: MANIATIS, T.  
;; AUTHORS: FRITSCH, E.F.  
;; AUTHORS: SAMBROOK, J.  
;; TITLE: MOLECULAR CLONING

;; JOURNAL: COLD SPRING HARBOR, NY  
;; VOLUME:  
;; ISSUE:  
;; PAGES:  
;; DATE: 1982  
;; DOCUMENT NUMBER:  
;; FILING DATE:  
;; PUBLICATION DATE:  
;; RELEVANT RESIDUES IN SEQ ID NO: 15:  
;; PUBLICATION INFORMATION:  
;; AUTHORS: SMITH, D.B.  
;; AUTHORS: JOHNSON, K.S.  
;; TITLE: SINGLE STEP PURIFICATION OF POLYPEPTIDES  
;; TITLE: EXPRESSED IN ESCHERICHIA COLI AS FUSIONS WITH  
;; TITLE: GLUTATHIONE S. TRANSFERASE  
;; JOURNAL: GENE  
;; VOLUME:  
;; ISSUE: 67  
;; PAGES: 31 - 40  
;; DATE: 1988  
;; DOCUMENT NUMBER:  
;; FILING DATE:  
;; PUBLICATION DATE:  
;; RELEVANT RESIDUES IN SEQ ID NO: 17:  
;; US-08-856-841-17  
;;  
Query Match 5.5%; Score 178.5; DB 4; length 387;  
Best Local Similarity 31.6%; Pred. No. 1.2e-07;  
Matches 48; Conservative 17; Mismatches 76; Indels 11; Gaps 2;  
;;  
QY 51 LGPPATPAVGPPTMASGGGAPMADNNEGADVGNASGNNHCDSTWLGDRVITSTRW 110  
DB 221 LPEVPAYNASEKYPMSVNSAE-ASTGAGGGGNSVKSMWSEGATFSANSVCTFSRPF 279  
QY 111 ALPTYNMHLKQISSASTGASND-----NHYFGYSTPWGYFDENRFHCHFSRPRW 160  
DB 280 LIPYDEHHYKVFSPAASSCHNASGKEAKVCTISPIMGYSTPWRYLDENALNLFSPLEF 339  
QY 161 QRLNNNWGFRPKRLNFKLFNIQVKEVTTNDG 192  
DB 340 QHLIENYGSIALDALTVTITSEIAVKDVTDKTG 371  
;;  
RESULT 15  
US-08-469-202-27  
; Sequence 27, Application US/08469202  
; Patent No. 5750875  
;; GENERAL INFORMATION:  
;; APPLICANT: STALKER, DAVID  
;; TITLE OF INVENTION: GLYCOGEN BIOSYNTHETIC ENZYMES IN PLANTS  
;; NUMBER OF SEQUENCES: 28  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: CALGENE, INC.  
;; STREET: 1920 FIFTH STREET  
;; CITY: DAVIS  
;; STATE: CA  
;; COUNTRY: USA  
;; ZIP: 95616  
;;  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB  
;; COMPUTER: Apple Macintosh  
;; OPERATING SYSTEM: Macintosh 7.1  
;; SOFTWARE: Microsoft word 5.1 (a)  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/469,202  
;; FILING DATE: 6-JUNE-95  
;; CLASSIFICATION: 800  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/016,881  
;; FILING DATE: 11\_FEB\_1993  
;; CLASSIFICATION: 800  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Elizabeth Lassen



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 2003, 18:34:40 ; Search time 28.8443 Seconds  
(without alignments)  
2466.245 Million cell updates/sec

Title: US-09-807-802A-15  
Perfect score: 3251  
Sequence: 1 TAPGKKRPVEQSPQEPDSSS.....NNGLYTEPRPIGTRYLTRPL 599

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep1:\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep2:\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep3:\*  
13: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2764.5	85.0	735	US-10-038-972A-13	Sequence 13, Appl
2	2764.5	85.0	735	US-10-293-478-1	Sequence 1, Appli
3	2759.5	84.9	598	US-10-038-972A-14	Sequence 14, Appl
4	2481.5	76.3	533	US-10-038-972A-15	Sequence 15, Appl
5	772	23.7	756	US-10-205-942-4	Sequence 4, Appli
6	118	3.6	2344	US-09-815-242-12713	Sequence 12713, A
7	116.5	3.6	5795	US-09-815-242-12610	Sequence 12610, A
8	108	3.3	1463	US-09-971-536-69	Sequence 69, Appl
9	106.5	3.3	1848	US-09-839-996-6	Sequence 6, Appli
10	106.5	3.3	1848	US-10-080-505-6	Sequence 6, Appli
11	106.5	3.3	2364	US-10-156-761-7834	Sequence 7834, Ap
12	106	3.3	420	US-10-062-254-254	Sequence 254, App
13	104.5	3.2	717	US-10-046-583A-1	Sequence 1, Appli
14	103.5	3.2	1016	US-09-815-242-5845	Sequence 5845, Ap
15	103.5	3.2	1265	US-10-198-070-69	Sequence 69, Appl
16	102	3.1	3712	US-10-108-605-103	Sequence 103, App

17	101.5	3.1	439	9	US-09-836-353A-145	Sequence 145, App
18	101.5	3.1	439	12	US-09-984-130-145	Sequence 145, App
19	101.5	3.1	1232	10	US-09-801-574-46	Sequence 46, Appl
20	101	3.1	676	11	US-09-801-368-302	Sequence 302, App
21	101	3.1	1411	15	US-10-080-505-17	Sequence 17, Appl
22	100.5	3.1	1714	15	US-10-098-916A-11	Sequence 11, Appl
23	100.5	3.1	4019	11	US-09-738-973-425	Sequence 425, App
24	100.5	3.1	4019	11	US-09-854-133-425	Sequence 425, App
25	100.5	3.1	4019	15	US-10-144-649A-425	Sequence 425, App
26	100	3.1	847	15	US-10-112-527-4	Sequence 4, Appli
27	99	3.0	623	15	US-10-108-605-125	Sequence 125, App
28	99	3.0	623	15	US-10-108-605-129	Sequence 129, App
29	99	3.0	824	15	US-10-046-583A-3	Sequence 3, Appli
30	99	3.0	1228	12	US-09-917-384-1	Sequence 1, Appli
31	99	3.0	1228	12	US-09-917-383-1	Sequence 1, Appli
32	98.5	3.0	515	15	US-10-097-340-212	Sequence 212, App
33	98.5	3.0	515	15	US-10-171-311-156	Sequence 156, App
34	98.5	3.0	620	12	US-09-931-009A-1	Sequence 1, Appli
35	98.5	3.0	644	12	US-09-919-039-121	Sequence 121, App
36	98.5	3.0	1000	15	US-10-055-412B-30	Sequence 30, Appl
37	98.5	3.0	1036	11	US-09-842-256-3	Sequence 3, Appli
38	98.5	3.0	1091	11	US-09-280-197-2	Sequence 2, Appli
39	98.5	3.0	1091	11	US-09-423-126-2	Sequence 2, Appli
40	98.5	3.0	2478	10	US-09-815-242-5816	Sequence 5816, Ap
41	98.5	3.0	2478	10	US-09-815-242-12967	Sequence 12967, A
42	98	3.0	425	15	US-10-156-761-12331	Sequence 12331, A
43	98	3.0	1770	10	US-09-841-132-444	Sequence 444, App
44	97.5	3.0	685	10	US-09-745-763-11	Sequence 11, Appl
45	97	3.0	659	15	US-10-046-583A-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1  
US-10-038-972A-13  
; Sequence 13, Application US/10038972A  
; Publication No. US20020192823A1  
; GENERAL INFORMATION:  
; APPLICANT: J. Bartlett  
; TITLE OF INVENTION: AAV VECTORS AND METHODS  
; FILE REFERENCE: 28335/36996US  
; CURRENT APPLICATION NUMBER: US/10/038, 972A  
; CURRENT FILING DATE: 2002-01-04  
; PRIOR APPLICATION NUMBER: US 60/260, 124  
; PRIOR FILING DATE: 2001-01-05  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 13  
; LENGTH: 735  
; TYPE: PRT  
; ORGANISM: adeno-associated virus 2 VP1 capsid protien  
; US-10-038-972A-13

Query Match	Best Local Similarity	Score	Pred. No.	DB	Length	Matches	Conservative	Mismatches	Indels	Gaps
QY	1	85.0%	82.8%	15	735	496	43	59	1	1
QY	1	TAPGKKRPVEQSPQEPDSSGIGTKGQPAKRLNFGQTDSESVDPDPLGEPATPAA	60							
Db	138	TAPGKKRPVEHSPVEPDSSSGTGAQGPARKRLNFGQTDADSVDPDPLGQPPAPSG	197							
QY	61	VGPTMASGGGAPMADNNEGADGVNAGSNMHCDSITWLGDRVITTTSTRTALPTYNNHLY	120							
Db	198	LGTNTMATGSGAPMADNNEGADGVNAGSNMHCDSITWLGDRVITTTSTRTALPTYNNHLY	257							
QY	121	KQISSASTGASNDNHFGYSTPWGYFDNRFHCHFSPRDWORLLINNNGFRPKRLNFKLF	180							
Db	258	KQISSQS-GASNDNHFGYSTPWGYFDNRFHCHFSPRDWORLLINNNGFRPKRLNFKLF	316							
QY	181	NIQVEVTNDGVTIANNTLSTVQVFSDEYQLPYVLGSAHQCLPPFPADVFMIPQYG	240							
Db	317	NIQVEVTQNDGTTIANNTLSTVQVFTDSEYQLPYVLGSAHQCLPPFPADVFMIPQYG	376							



QY	241	YLTINNGSOAVGRSSFYCLEYFPSPÖMLRTGNNTFTSYTFEEVFPFHSSYAHSQSLDRIMNP	300
Db	377	YLTINNGSOAVGRSSFYCLEYFPSPÖMLRTGNNTFTSYTFEEDVPFHSSYAHSQSLDRIMNP	436
QY	301	LIDÖLYYYLNRTQÖNGSSAQÖNKDLLFSRGSFAGNSVQPKNWLPGPCYRQÖRVSKITDNN	360
Db	437	LIDÖLYYYLSRTNTPSGTTQÖSRLQFSQAGASDIRQÖSRNWLPGPCYRQÖRVSKITSADNN	496
QY	361	NSNFTWTGASKYNLNGRESIINPGTAMASHKODEKFEPPMSGVMI FGESAGASNTALDN	420
Db	497	NSFSWTGATKYHLNGRDSLVPNGPAMASHKODEKEFPQÖGVLIFGQÖSEKTNVDIEK	556
QY	421	VMTDDEEIKATNPVATERFGTVAVNQÖSSSTDPAATGDVHAMGALPGMWQÖDRDYYLQÖP	480
Db	557	VMTDDEEIRTNPNVATEQÖGSVSSTNLQÖRNGQAATADVNTQÖGVLPGMWQÖDRDYYLQÖP	616
QY	481	IMAKIPTHDDGHFHPSPLMGGFGLKNPFPÖILLKNTPVPANPFAEFSAATKFAFITQÖSTG	540
Db	617	IMAKIPTHDDGHFHPSPLMGGFGLKHPFPÖILLKNTPVPANPSTTFSAAKFAFITQÖSTG	676
QY	541	QVSVEIEMELÖKENSKRWNPEVQÖTSSNYAKSANDFTVDNNGLYTEPRPIGTRYLTRPL	599
Db	677	QVSVEIEMELÖKENSKRWNPEIQÖTSSNYAKSANDFTVDITNGVYSEPRPIGTRYLTRPL	735

```

RESULT 2
US-10-293-478-1
; Sequence 1, Application US/10293478
; Publication No. US20030078411A1
; GENERAL INFORMATION:
; APPLICANT: PATEL, SALIL D.
; APPLICANT: MCARTHUR, JAMES G.
; TITLE OF INVENTION: ADENO-ASSOCIATED VIRUS CAPSID IMMUNOLOGIC DETERMINANTS
; FILE REFERENCE: 39672
; CURRENT APPLICATION NUMBER: US/10/293,478
; CURRENT FILING DATE: 2002-11-14
; PRIOR APPLICATION NUMBER: US/09/321,589
; PRIOR FILING DATE: 1999-05-28
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 735
; TYPE: PRT
; ORGANISM: Adeno-associated virus
US-10-293-478-1

```

Query Match	85.0%;	Score 2764.5;	DB 15;	Length 735;
Best Local Similarity	82.8%;	Pred. No. 4.3e-23;		
Matches 496;	Conservative 43;	Mismatches 59;	Indels 1;	Gaps 1
Qy	1	TAPGKKRPVEQSPQEPDSSSGTGTGQOPAKKRLNFGQTGDSIESVPDQPLGEPPATPA	60	
Db	138	TAPGKKRPVEHSPVEPDDSSSGTGAKGQOPAKKRLNFGQTGDADSVEDPQPLGQPPAAPSG	197	
Qy	61	VGPPTMASGGGAPMADNNEGADGVENASGNWHCDSTWLGDVRVITSTRTALPTYNNHLY	120	
Db	198	LGTINTMATGSGAPMADNNEGADGVENSSGNWHCDSTWMDRVITSTRTALPTYNNHLY	257	
Qy	121	KQISSASTGASNDNHFGYSTPWGYFDNRFHCHFSPRDWQRLINNNGFRPKRLNFKLF	180	
Db	258	KQISSQS-GASNDNHFGYSTPWGYFDNRFHCHFSPRDWQRLINNNGFRPKRLNFKLF	316	
Qy	181	NIQVEVTNDGVTTIANNLTSTVOVFSDSSEYQLPYVLGSAHQGLPPPADVEMIPQYG	240	
Db	317	NIQVEVTQNDGTTIANNLSTVQVFTDSEYQLPYVLGSAHQGLPPPADVEMVPQYG	376	
Qy	241	YLTNNGSQAVGRSSFYCLEYFSPQMLRTGNFTFSYTFEEV/PFHSSYAHQSQSLDRLMNP	300	
Db	377	YLTNNGSQAVGRSSFYCLEYFSPQMLRTGNFTFSYTFEDV/PFHSSYAHQSQSLDRLMNP	436	
Qy	301	LIDQYLYLNRQTQNSGAQNKDLLFSRGSPPAGMSVQPKNWLPGPCYRQORVSKTKTDNN	360	

Db	437	LIDÖYLYYLSRINTPSCGTTQTSRLOFSQAGASDIRDQSRNWLPGPCYRQORVSKTSADNN	496
Qy	361	NSNFTWTGASKYNLNGRESITNPGTAMASHKDEDEKFFPMGVMIPGKESAGASNTALDN	420
Db	497	NSEYSWTGATKYHLNGRDSLVPNGPAMASHKDEEKFEPQSGVLI FGKQSEKTNVDIEK	556
Qy	421	VMITDEEIKATNPVATEREGTYAVNFQSSSTDTPATGVDVHMGALPGMWODRDVYLOGP	480
Db	557	VMITDEEEIRTNPVATEBQYGSVSTNLQRGNRQAAATADVNTQGVLPGMWODRDVYLOGP	616
Qy	481	IWAKIPHTDGHFHPSPLMGGGGLKNPPÖILIKNTPV PANP PAEFSATKASFITÖYSTG	540
Db	617	IWAKIPHTDGHFHPSPLMGGGGLKHPÖÖILIKNTPV PANP STFSAKFASFITÖYSTG	676
Qy	541	QVSVEIEMELÖKENS KRWNPEVOYTSNYAKSANDFTVDNNGLYTEPRP IGT RYLT RPL	599
Db	677	QVSVEIEMELÖKENS KRWNPEÖYTSNYNKS VNVDFTVDTNGVYSEPRP IGT RYLT RNL	735

```

RESULT 3
US-10-038-972A-14
; Sequence 14, Application US/10038972A
; Publication No. US20020192823A1
; GENERAL INFORMATION:
; APPLICANT: J. Bartlett
; TITLE OF INVENTION: AAV VECTORS AND METHODS
; FILE REFERENCE: 28335/36996US
; CURRENT APPLICATION NUMBER: US/10/038, 972A
; CURRENT FILING DATE: 2002-01-04
; PRIOR APPLICATION NUMBER: US 60/260,124
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
; LENGTH: 598
; TYPE: PRT
; ORGANISM: adeno-associated virus 2 VP2 capsid protien
US-10-038-972A-14

```

	Query Match	84.9%;	Score 2759.5;	DB 15;	Length 598;	
	Best Local Similarity	82.8%;	Pred. No. 8.7e-235;			
	Matches 495;	Conservative 43;	Mismatches 59;	Indels 1;	Gaps 1;	
Qy	2	APGKKRPVEQSPOEPDSSSGIGKTGOOPAKKRLNFGQTGSSEVPDPQPLGEPPATPAAV	61			
Dp	2	APGKRPRVESHSPVEPDSSSGTGAGQQPARKRLNFGQTGDADSVDPQLGPAPAASGL	61			
Qy	62	GPTTMASGGGAPMADNNNEGADVGNASGNMHCDSTWLGDVRITTTSTRTALPTYNNHLYK	121			
Dp	62	GTNTMATGSGAPMADNNNEGADVGNSSGNMHCDSTWNGRIVITTSRTWALPTYNNHLYK	121			
Qy	122	QISSASTGASNDNHFYGYSTPWGYFDENFRFHCHFSPRDMORLINNNWGFRPKRLNFKLFN	181			
Dp	122	QISSQS-GASNDNHFYGYSTPWGYFDENFRFHCHFSPRDMORLINNNWGFRPKRLNFKLFN	180			
Qy	182	IQVEVTINDGVTTIANNLISTVQVFSDEYQLPYVLGSAHQCLPPFPADVFMIPQGY	241			
Dp	181	IQVEVTQNDGTTIANNLISTVQVFTDSEYQLPYVLGSAHQCLPPFPADVFMVPOGY	240			
Qy	242	LTLNNGSQAVGRSSFYCLEYFPSOMLRTGNNFTFSYTEEEVPEFHS SYAHOSQLDRIMNPL	301			
Dp	241	LTLNNGSQAVGRSSFYCLEYFPSOMLRTGNNFTFSYTEEDVPEFHS SYAHOSQLDRIMNPL	300			
Qy	302	IDQYLYYLNRTONOGSQAQNKDLLFSRGSPAGMSVQPKMLPGPCYRQQRVS KTKTDNNN	361			
Dp	301	IDQYLYYLSRTNTPSGITTQSRLOFSQAGASDIRDQSRMWLPGPCYRQQRVSKTSADNNN	360			
Qy	362	SNFTWTGASKYNLNGRESIINPGTAMASHKDEDKFFPMSCVMIFEKESAGASNTALDNV	421			
Dp	361	SEYSWTGATKYHLNGRDSLVPNGPAMASHKDEEKFPOSQVLIIFGKGSEKTNVDIEKV	420			
Qy	422	MITDEEIKAINPVATERFGTVAVNFQSSSTDPAIDGHANGALPGMVWD RDRVYLGPI	481			

Db 421 MITDEEIRTNPVATEQYGSVSTNLQRGNQATADVNTQGVLPQMWQDRDVLQGP 480  
Qy 482 WAKIPHTDGHFSPPLMGFGGLKNPPQILIKNTVPANPAPFAFSATKPFASFITQYSTGQ 541  
Db 481 WAKIPHTDGHFSPPLMGFGGLKNPPQILIKNTVPANPAPSTTFSAKFASFITQYSTGQ 540  
Qy 542 VSVEIEMELOKENSKRWNPEVQYTSNVAKSANVDFVDNNGLYTEPRPIGTRYLTRPL 599  
Db 541 VSVEIEMELOKENSKRWNPEIQYTSNVAKSANVDFVDNNGLYTEPRPIGTRYLTRPL 598

## RESULT 4

US-10-038-972A-15  
; Sequence 15, Application US/10038972A  
; Publication No. US20020192823A1  
; GENERAL INFORMATION:  
; APPLICANT: J. Bartlett  
; TITLE OF INVENTION: AAV VECTORS AND METHODS  
; FILE REFERENCE: 28335/36996US  
; CURRENT APPLICATION NUMBER: US/10/038,972A  
; CURRENT FILING DATE: 2002-01-04  
; PRIOR APPLICATION NUMBER: US 60/260,124  
; PRIOR FILING DATE: 2001-01-05  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 15  
; LENGTH: 533  
; TYPE: PRT  
; ORGANISM: adeno-associated virus 2 VP3 capsid protien  
US-10-038-972A-15

Query Match 76.3%; Score 2481.5; DB 15; Length 533;  
Best Local Similarity 83.3%; Pred. No. 2.5e-210;  
Matches 445; Conservative 37; Mismatches 51; Indels 1; Gaps 1;

Qy 66 MASGGAPADNNEGADGVGNASGNWHDSTWLGDVITTTSTRVALPTYNHLYKOISS 125  
Db 1 MATGSGAPADNNEGADGVGNSSGNWHDSTWMDRVITTTSTRVALPTYNHLYKOISS 60  
Qy 126 ASTGASNDNHFFGYSTPWGYFDNFRFHCHFSPRDWQRLINNNGFPRKLNFKLFNIQVK 185  
Db 61 QS-GASNDNHFFGYSTPWGYFDNFRFHCHFSPRDWQRLINNNGFPRKLNFKLFNIQVK 119  
Qy 186 EYTTNDGVTTIANNLTSTVQVFSDEYQLPYLGSAGGCLPPPADVFMIPQYGYLTIN 245  
Db 120 EYTTNDGVTTIANNLTSTVQVFTDSEYQLPYLGSAGGCLPPPADVFMIPQYGYLTIN 179  
Qy 246 NGSQAVGSSFFCYCLEYFSPQMLRTGNNTFSYTFEEVPFHSSYAHQSOLDRLMNPIDQY 305  
Db 180 NGSQAVGSSFFCYCLEYFSPQMLRTGNNTFSYTFEDVPFHSSYAHQSOLDRLMNPIDQY 239  
Qy 306 LYYLNRTPQNSGSAQNKDLLFSRGSFAGMSVQPKNMLPGPCYRQORVSKTKTDNNNSNFT 365  
Db 240 LYYLNRTPQNSGSAQNKDLLFSRGSFAGMSVQPKNMLPGPCYRQORVSKTKTDNNNSNFT 299  
Qy 366 WTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMGSMIFGKESAGASNTALDNYMTD 425  
Db 300 WTGATKYHLNGRDSLVPNGPAMASHKDEDEKFFPQSGVLI FGKQSEKTNVDIEKVMITD 359  
Qy 426 EEEIKATNPVATERFGTAVANFQSSSTDPATGDVHAMGALPGMWQDRDVLQGPWAKI 485  
Db 360 EEEIRTNPVATEQYGSVSTNLQRGNQATADVNTQGVLPQMWQDRDVLQGPWAKI 419  
Qy 486 PHTDGHFSPPLMGFGGLKNPPQILIKNTVPANPAPFAFSATKPFASFITQYSTGQSV 545  
Db 420 PHTDGHFSPPLMGFGGLKNPPQILIKNTVPANPAPSTTFSAKFASFITQYSTGQSV 479  
Qy 546 IEMELOKENSKRWNPEVQYTSNVAKSANVDFVDNNGLYTEPRPIGTRYLTRPL 599  
Db 480 IEMELOKENSKRWNPEIQYTSNVAKSANVDFVDNNGLYTEPRPIGTRYLTRPL 533

## RESULT 5

US-10-205-942-4  
; Sequence 4, Application US/10205942  
; Publication No. US20030053990A1  
; GENERAL INFORMATION:  
; APPLICANT: University of No. US20030053990A1th Carolina-Chapel Hill  
; APPLICANT: Rabinowitz, Joseph E.  
; APPLICANT: Samulski, Richard J.  
; APPLICANT: Xiao, Weidong  
; TITLE OF INVENTION: VIRUS VECTORS AND METHOD OF MAKING AND ADMINISTERING THE SAME  
; FILE REFERENCE: 5470-186  
; CURRENT APPLICATION NUMBER: US/10/205,942  
; CURRENT FILING DATE: 2002-07-26  
; NUMBER OF SEQ ID NOS: 59  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 4  
; LENGTH: 756  
; TYPE: PRT  
; ORGANISM: Adeno-associated virus  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)-(2271)  
; OTHER INFORMATION: B19/AAV chimeric capsid coding sequence  
US-10-205-942-4

Query Match 23.7%; Score 772; DB 15; Length 756;  
Best Local Similarity 32.9%; Pred. No. 3.1e-59;  
Matches 195; Conservative 87; Mismatches 257; Indels 54; Gaps 16;

Qy 1 TAPGKKRPVEQSPQEPDSSGGIGTKGQOPAKKRLNFGQTDSESVDPDQPLGEPATPA 60  
Db 138 TAPGKKRPVEHSPVEPDSSGGIGTKGQOPAKKRLNFGQTDSDVPDQPLGQPAAPSG 197  
Qy 61 VGPITMASGGAPMADNNEGADGVGNASGNWHDSTWLGDVITTTSTRVALPTYNHLY 120  
Db 198 LGTNTWTSVNSAE-ASTGAGGGGNSVKSMSSEGATFSANSVTCTFSRQFLIPYDPEHHY 256  
Qy 121 KQISSASTGASND-----NHFFGYSTPWGYFDNFRFHCHFSPRDWQRLINNNGWF 170  
Db 257 KVFSPAASSCHNASGKEAKVCTISPIMGYSTPWRYLDFNALNLFSPLEFQHLIENYGS 316  
Qy 171 RPKRLNFKLFNIQVKEVT--TNDGVTTIANNLTSTVQVFSDEYQLPYLGSAGGCLPP 228  
Db 317 APDALVTITSEIAVKDVTDKTGGGV-QVTDSTGRCLMLVDHEKYPYVLGGQDITLAP 375  
Qy 229 FPADVFMIPQYGYLTIN-GSQAVG-----RSSFYCLEYFSPQMLRTGNNTFSYT 278  
Db 376 LPINWYFPQYALTLTGVDVNTQGISGDSKKLASEESAFVLEHSSFOLLGTGTATMSYK 435  
Qy 279 FEEVPFHSSYAHQSOLDRLMNPIDQYLYLNRTPQNSGSAQNKDLLFSRGSFAGMSVQ 338  
Db 436 FPPVPEPNEGCSQHFEYENPFL--YGSRLGVPDITLGGDPKFRSL-----THEDHAI 487  
Qy 339 KMWLPGPCYRQORVSKTKTDNNNSNFTWTGASKYNLNGRESIINPG-TAMASHKDEDEK 397  
Db 488 QNFMGPVLVNSVSTKEGDSNTGAGKALTGLSTGTSQNTIRISLRGPVSQPYHHMDTDKY 547  
Qy 398 FPMGSMIFGKESAGASNTALDNV-----MITDEEIKATNPVATERFGTAVANFQSS 450  
Db 548 VTGINAISHGQTTYG--NAEKEYQGVGRFPNEKEQLKOLQGLNMHTY-----FPMK 598  
Qy 451 STDPATGDVHAMGALPGMWQDRDVLQGPWAKIPHTDGHFSP-PLMGFGGLKNPPQ 509  
Db 599 GTQYTDQIE-RPLMVGSVNNRRALAHESQLMSKIPNLDSFKTQFALGGWGLHQPPQ 657  
Qy 510 ILIKNTVPANPAPFAFSATKPFASFITQYSTGQSVVEIEMEL-QKENSKRWNPE 561  
Db 658 IFLK--ILPQSGPIGIGIKSMGITTLVQYAVGIMTVMTFKLGPBKATGRWNPO 708

## RESULT 6

US-09-815-242-12713  
; Sequence 12713, Application US/09815242  
; Patent No. US20020061569A1

GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
PRIOR FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 12713  
LENGTH: 2344  
TYPE: PRT  
ORGANISM: Staphylococcus aureus  
US-09-815-242-12713

Query Match 3.6%; Score 118; DB 10; Length 2344;  
Best Local Similarity 20.4%; Pred. No. 0.89;  
Matches 87; Conservative 57; Mismatches 182; Indels 100; Gaps 18;

QY 37 GQTGDESEVP-DPQPLGEPATPAV-----PTTASGGGAPMA 75  
DB 326 GNGGDGIGFAFSPVLGETGLNGAAGVIGLSNAGFKLDYHNTSTPNSAKAKADPSN 385  
QY 76 DNEGADGVGNASGNWHDSTWLGDRTTSTRFVALPTYNHLYKQISSASTGASNDNH 135  
DB 386 VAGGAGFAGFVTDSTYGASTYSSSTADNAKLVQPTNN-----TFQDFDIN 434  
QY 136 YFGYSTPWGYFDENRFCHFSRWDQRLINNMGFRPKRLNFKLNIQVKEVTTNDGVT 195  
DB 435 YNG-----DTKVMYKYAGQWTNRNI-SDWIAKSGTTNPSL-----SMTASTGATN 480  
QY 196 IANNLTSTVQVFSDEYQLPYVLGSAHQGLPP--FPADVEMIPQYGYLTINNGSQAVGR 253  
DB 481 LQOVQGTFEYTESAVTQVRYVDVTGKDIIPKTYSGNVQV-----VTIDNQSALTA 535  
QY 254 SSF-----YCLEYFPS---QMLRTGNNFTFSYTFEEVFPFHSSYAHSQSLDRLMNP 301  
DB 536 KGYNYSVDSSTYASTYDINKTKMTNAGQSVTYFTDVKAPTIVTGNQTEVGKTMNPI 595  
QY 302 IDQYLYLNRTONOGSSAQNKDLFSRGSAGMSV-QPKNWLPGPCYR--QQRVSKTKTD 358  
DB 596 V-----LTTTNGTGTNT---VTGLPSGLSYDSATNSLIGTPKIGQSTVTVSTD 645  
QY 359 --NNNSFTWTGASKYNLNGRESLIN-----PGTAMASHKDDDEKFFPMGVMIFGES 410  
DB 646 QANNKSTTFT-----INVVDTPAPYTPIGDKSE-VFSPISINIAQTODN 691  
QY 411 AGASNT 416  
DB 692 SGNNAV 697

RESULT 7

US-09-815-242-12610  
Sequence 12610, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
PRIOR FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 12610  
LENGTH: 5795  
TYPE: PRT  
ORGANISM: Staphylococcus aureus  
US-09-815-242-12610

Query Match 3.6%; Score 116.5; DB 10; Length 5795;  
Best Local Similarity 19.1%; Pred. No. 4.6;  
Matches 128; Conservative 86; Mismatches 250; Indels 207; Gaps 35;

QY 40 GDSESVDPQPLGEPATPAVGPPTTASGGGAPMADNEGADGVGNASGNWHDSTWL 99  
DB 1696 GNGSSADYFKLSNGSAIPDA---TITWVGQAPNKDNTRIGEDI-NVTANILID---G 1747  
QY 100 DRVITSTRFVALPTYNHLYKQISSASTGASNDNHFGYSTPWGYFDENRFCH--- 154  
DB 1748 ETTPITKTAT-----YKVS-----SVKGVFETNRGAVFPGVSD 1782  
QY 155 -FSRWDQRLINNMGFRPKRLNFKLNI-----QVKEVTTNDGVTTIANNLTSTVQVFS 208  
DB 1783 VYDAKQYKPVNDSTQNAQRNMFQFTNSYGPSKDVGISTRDIRVTYDNHQTIKILA 1842  
QY 209 DSEYQLPYVLGSA-----HQGLPPFPADVEMIPQYGYLTINNGSQ 250  
DB 1843 KVKPPDPRIIDGNSVTYKAGLTNQIKINNVLSSSSIKLFKAD--NTP---LTTTNTY 1896  
QY 251 VGRSSFYCL-EYFPSQMLRTGNNFTFSYTFEEVFPFHSSYAHSQSLDRLMNP 309  
DB 1897 SGNNAVTVSDALPNGVIKARS---SITMNNVTYTTQDEHGRAIDVTRNESVD----- 1946  
QY 310 NRTQNGSS-----AQNKDLFSRGS-----PAGMSV-----QPKNW- 341  
DB 1947 ---SNDGATVTVTPLOQATTEGAVFIKGGDGFDFGHERFIQNPBGATVAMHDNPD 2003  
QY 342 -----LPGCYRQQRVSKTKTDNNNSFTWTGASKYNLNGR 377  
DB 2004 NTVGNTHTKTAVTVLPSSGQTRNVEVVKVYPVANAKAPSRDYKQONLT-----NGT 2054  
QY 378 EST-----INPGT-----AMASHKDDDEKFFPMGVMIFGESAGASNTALDNVMT-- 424



Db 2055 DAINYTFDENTNTNGITAAWANNRQPPNNQ---QAGVQHLNVDTYTPGITAARVPVTN 2111  
Qy 425 -DEEEKATNPVATERFGTVAVNFQSSSTDPATGVDHAMGA--LP--GMV--WQDRDVL 477  
Db 2112 VYQFEFPQTSYTTTVG-GTILANGTQ-----ASGYAHMQNANGIPTDGYTKWNNATGT 2164  
Qy 478 QGPIWAKI--PHT-----DGHFHPSPLMGFGLKN--PPQILIKNTPVPA--- 519  
Db 2165 NDANWAMNKNPAKVNNAKYDVIXNGHTPATSLPAKFVVKDVPAPKPTVETAGAITI 2224  
Qy 520 NPPAFESATKPFASFITQYS-----TGQVSVEIEMELQKENSKRWNPEVQYTSNYAKSA 572  
Db 2225 TPGANQTVNTHAGNVTTYADKLVIKRNQNV---VTIFTRRNNTSPWVKE---ASAATVA 2277  
Qy 573 NVDFTVDNNGI 583  
Db 2278 GIAGT--NNGI 2286

RESULT 8

US-09-971-536-69  
; Sequence 69, Application US/09971536  
; Patent No. US20020159976A1  
; GENERAL INFORMATION:  
; APPLICANT: Glenn, Matthew  
; APPLICANT: Havukkala, Ilkka  
; APPLICANT: Bloksberg, Leonard  
; APPLICANT: Lubbers, Mark  
; APPLICANT: Dekker, James  
; APPLICANT: Christensson, Anna  
; APPLICANT: Holland, Ross  
; APPLICANT: O'Toole, Paul  
; APPLICANT: Reid, Julian  
; APPLICANT: Coolbear, Timothy  
; TITLE OF INVENTION: Lactobacillus rhamnosus Polynucleotides, Polypeptides and Methods  
; TITLE OF INVENTION: Using Them  
; FILE REFERENCE: 1043c2  
; CURRENT APPLICATION NUMBER: US/09/971,536  
; PRIOR APPLICATION NUMBER: 2001-10-02  
; PRIOR FILING DATE: 2000-08-08  
; PRIOR APPLICATION NUMBER: U.S. No. US20020159976A1 09/634,238  
; PRIOR FILING DATE: 2000-11-28  
; PRIOR APPLICATION NUMBER: PCT/NZ01/00160  
; PRIOR FILING DATE: 2001-08-08  
; NUMBER OF SEQ ID NOS: 83  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 69  
; LENGTH: 1463  
; TYPE: PRT  
; ORGANISM: Lactobacillus rhamnosus  
US-09-971-536-69

Query Match 3.3%; Score 108; DB 11; Length 1463;  
Best Local Similarity 21.4%; Pred. No. 3.4;  
Matches 114; Conservative 67; Mismatches 214; Indels 138; Gaps 27;

Qy 1 TAPGKRPRVEQSPQEPDSSSGIGKTGQPAKKRLNFGQTGDSSEVPDPQPLGEPATPA 60  
Db 85 TSSGSSDTSQANEGNSASVTSASANSASATNTSEGQ-----PVNEAVSNDAS 134  
Qy 61 VGPTTMASGGGAPMADNNEGADGVNAGSNWCHDSTWLGDRVITTSRTWALPTYNNH-- 118  
Db 135 SADVSTASEFDAMADSTV---SVINVQSDFVMDVS--GDRQ-SYAYRPNLIINGNNHTI 188  
Qy 119 -LYKQISSASTGAS-----NDNHYFGYSTPWGYPDFNRFCHFSPRDWORLNNNMGF 170  
Db 189 DFQKKYHEADPTSSQNESFTINDLNNYGS-----W-----WG- 221  
Qy 171 RPKRLNFKLNIQVKEVTNDGV--TTIANNLSTVQVFSDEYQLPYVLGSAHQGLPP 228  
Db 222 -----PVTIKGSKPKDGDHGSVFNNTVTYGAQLMWGYITKAFIKGNT----- 264

Qy 229 FPADVFMIPQYQ-YLTINNGS-----QAVGRSSFYCLBYRPSQMLRTGNMFTFSYTFEEVP 283  
Db 265 -----KIQSVGSYVSPLDGSIQTQTQGLGNQONFOISYLEVLPGATYTTGTTGNTVEYV 318  
Qy 284 FHSSYA-----HQSILRLMNPILIDQLYLYLNFTQNGSSAQNKDLFSRGS 330  
Db 319 DGGSPFVDKATVNLQRTDASKSNERGTNALID-----TQGNVEFKD-----GS 363  
Qy 331 PAGMSVQPKMWLP---GPCYRQQRVSKTKTDNNNSNFTW--TGASKYNLNGRESI--INPG. 384  
Db 364 TV---ILNKNALVKDGFAPYIEDGGLTVDKNATVSTGATIGNIPVRIDGTGVNLNEG 420  
Qy 385 TAMASHKDEDEKFFPMSGVMIFGKESAG---ASNTALD-NVMITDEEEKATNPVATER 439  
Db 421 ----SHMTITQNGAPKLGYGFINIKGTGGFFVASGSTLDLNVGTGTSVNAIN-VANDG 475  
Qy 440 FGTVAVNFQSSST-DPATGVDHAMGALPGMWQDR--DVILOGPWAKIPHTD 489  
Db 476 QLSFAQDATANLTIDGTEGAHLL-----KVGDDANINIMPKSVLFKITDND 523

RESULT 9

US-09-839-996-6  
; Sequence 6, Application US/09839996  
; Publication No. US20030009010A1  
; GENERAL INFORMATION:  
; APPLICANT: St. Geme III, Joseph W.  
; TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/839,996  
; FILING DATE: 20-Apr-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/296,791  
; FILING DATE: 25-AUG-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Treacartin, Richard F.  
; REGISTRATION NUMBER: 31,801  
; REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 761-1989  
; TELEFAX: (415) 398-3249  
; TELE: 910 277299  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1848 amino acids  
; TYPE: amino acid  
; TOPOLOGY: unknown  
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:  
US-09-839-996-6

Query Match 3.3%; Score 106.5; DB 12; Length 1848;  
Best Local Similarity 18.7%; Pred. No. 6.4;  
Matches 133; Conservative 79; Mismatches 264; Indels 237; Gaps 31;

Qy 60 AVGPPTMASGGGAPMA-DNNEGAD-----GVGNASGNWCHD----- 94

```
Db 363 ATGSTITGGEPPLSVLTDGKDPNHGKSYTLKSGTLLTNHIDQAGGLFEEGDYE 422
Qy 95 -----STWLGDRVITT--STRTVAL--PTY----- 115
Db 423 VKGTSdstwkgagvsVADGKTvtwkvnhPKYDLAKIGKGLVVEGKNEGGLKVGDG 482
Qy 116 -----NNHL--YKQ---ISSASTGASNDNH-----YFGYSTPMGYFDENRFHC 153
Db 483 TVILKQKADANNKVQAFSQVGIVSGRSTLVLDNDKQVDPNSTYFGFR--GGRLDLNGSL 540
Qy 154 HF----SPRDWQRLINNNWGFRRPKRLNFKLFNIQVKEVTTNDGVTTLANNLTSTVQVFS 209
Db 541 TFDHIRNIDDGARVNVNHN-----TNTSNITI---TGESLITNPNTITTSYNIEAD 588
Qy 210 SEYQLPYVLGSAHQGLPPFPADVFMIPQYGYTLTNNGSQ-----VG 252
Db 589 DDH--PLRIRSI-----PYROLYFNQDNRSYTLKKGASTR3SELPONSGESNENWLYMG 640
Qy 253 RSSFYCLEYFPQSOMLRTGNFTFSYTFEEVPHSSYAHQSGLDRLMNPPLIDQYLYLNRT 312
Db 641 RTSDAARNVNMHNINNERMNGFNGYFGEE---ETKATQNGKLNVTFNKSDQNRFLLTGG 697
Qy 313 QNOSGSAQ-NKDLFSRGSF-----AGMSVQPKNWLPGPCYRQQRVSKTKTDNNNSNF 364
Db 698 TNLNGDLNVEKGTFLSGRPTPHARDIAGISSTKCD---PHFTENNEVVEDDWINRNF 753
Qy 365 TWTGASKYNLNGRESIIN-----PGTAMASHKDEDEKFFPMGVMIFGKESAG----- 412
Db 754 K---ATTMNVGTGNASLYSGRNVANITSNITASNAQVHIGYKTDIVCVRSYDTGYVTCH 810
Qy 413 ---ASNTALD-----NMITDEEEI--KATNPVATERFGTVAVNFQSSSTDPATG 457
Db 811 NSNLSEKALNSFNPTNLRGVNLTENASFTLGKANLFGTIGISQVNLKENSHPMLTG 870
Qy 458 DVHAMGALPGMWQDRDVLQGPWAKIPHTDGHFHPSPLMGFGGLKNPPQILIKNTPV 517
Db 871 NSNV-----NQLNLNGHIH-----LNAQNDANKVTTYNTLT 902
Qy 518 PANPPEFSATKFAF-----ITQYSTQGVSEIEWELOKENSKRMNPEVOYTSNY 568
Db 903 VNSLGSNGSFYVWDFTNKSNKVVNKSATGNFTLQV--ADKTGEPNHNETLTFDASN 959
Qy 569 AKSANVDFTVDN-----NGLYTEPRP-----IGTRYLTRP 598
Db 960 ATRNNLEVTLANGSVDRGAMWYKLRNVNGRYDLYNPEVEKRNQTVDTTNTTTP 1012
```

## RESULT 10

```
US-10-080-505-6
; Sequence 6, Application US/10080505
; Publication No. US20030073166A1
; GENERAL INFORMATION:
; APPLICANT: St. Geme, Joseph W.
; TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS
; FILE REFERENCE: A-59941-1/RFT/DCF/DHR
; CURRENT APPLICATION NUMBER: US/10/080,505
; CURRENT FILING DATE: 2002-02-22
; PRIOR APPLICATION NUMBER: US 08/296,791
; PRIOR FILING DATE: 1994-10-25
; PRIOR APPLICATION NUMBER: US 09/839,996
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 1848
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-10-080-505-6
```

Query Match 3.3%; Score 106.5; DB 15; Length 1848;  
Best Local Similarity 18.7%; Pred. No. 6.4;  
Matches 133; Conservative 79; Mismatches 264; Indels 237; Gaps 31;

```
Qy 60 AVGPTTMAAGGAPMA-DNNEGAD-----GVGNAGNWHCD----- 94
Db 363 ATGSTITGGEPPLSVLTDGKDPNHGKSYTLKSGTLLTNHIDQAGGLFEEGDYE 422
Qy 95 -----STWLGDRVITT--STRTVAL--PTY----- 115
Db 423 VKGTSdstwkgagvsVADGKTvtwkvnhPKYDLAKIGKGLVVEGKNEGGLKVGDG 482
Qy 116 -----NNHL--YKQ---ISSASTGASNDNH-----YFGYSTPMGYFDENRFHC 153
Db 483 TVILKQKADANNKVQAFSQVGIVSGRSTLVLDNDKQVDPNSTYFGFR--GGRLDLNGSL 540
Qy 154 HF----SPRDWQRLINNNWGFRRPKRLNFKLFNIQVKEVTTNDGVTTLANNLTSTVQVFS 209
Db 541 TFDHIRNIDDGARVNVNHN-----TNTSNITI---TGESLITNPNTITTSYNIEAD 588
Qy 210 SEYQLPYVLGSAHQGLPPFPADVFMIPQYGYTLTNNGSQ-----VG 252
Db 589 DDH--PLRIRSI-----PYROLYFNQDNRSYTLKKGASTR3SELPONSGESNENWLYMG 640
Qy 253 RSSFYCLEYFPQSOMLRTGNFTFSYTFEEVPHSSYAHQSGLDRLMNPPLIDQYLYLNRT 312
Db 641 RTSDAARNVNMHNINNERMNGFNGYFGEE---ETKATQNGKLNVTFNKSDQNRFLLTGG 697
Qy 313 QNOSGSAQ-NKDLFSRGSF-----AGMSVQPKNWLPGPCYRQQRVSKTKTDNNNSNF 364
Db 698 TNLNGDLNVEKGTFLSGRPTPHARDIAGISSTKCD---PHFTENNEVVEDDWINRNF 753
Qy 365 TWTGASKYNLNGRESIIN-----PGTAMASHKDEDEKFFPMGVMIFGKESAG----- 412
Db 754 K---ATTMNVGTGNASLYSGRNVANITSNITASNAQVHIGYKTDIVCVRSYDTGYVTCH 810
Qy 413 ---ASNTALD-----NMITDEEEI--KATNPVATERFGTVAVNFQSSSTDPATG 457
Db 811 NSNLSEKALNSFNPTNLRGVNLTENASFTLGKANLFGTIGISQVNLKENSHPMLTG 870
Qy 458 DVHAMGALPGMWQDRDVLQGPWAKIPHTDGHFHPSPLMGFGGLKNPPQILIKNTPV 517
Db 871 NSNV-----NQLNLNGHIH-----LNAQNDANKVTTYNTLT 902
Qy 518 PANPPEFSATKFAF-----ITQYSTQGVSEIEWELOKENSKRMNPEVOYTSNY 568
Db 903 VNSLGSNGSFYVWDFTNKSNKVVNKSATGNFTLQV--ADKTGEPNHNETLTFDASN 959
Qy 569 AKSANVDFTVDN-----NGLYTEPRP-----IGTRYLTRP 598
Db 960 ATRNNLEVTLANGSVDRGAMWYKLRNVNGRYDLYNPEVEKRNQTVDTTNTTTP 1012
```

## RESULT 11

```
US-10-156-761-7834
; Sequence 7834, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 7834
; LENGTH: 2364
; TYPE: PRT
```

ORGANISM: Streptomyces avermitilis  
US-10-156-761-7834

Query Match 3.3%; Score 106.5; DB 15; Length 2364;  
Best Local Similarity 20.5%; Pred. No. 9.3;  
Matches 96; Conservative 58; Mismatches 158; Indels 157; Gaps 22;

QY 14 QEPDSSGIGKTGQOPAKRLNFGQTSSESVDPQPLGEPATPAVPTTMASGGAP 73  
DB 849 QEVQVTTGAG-AGQKTKKSWLYRGLSGDRTSKADAS-----ATKVT----- 889  
QY 74 MADNNEGADVGNASGNWCHDSTWLGDRVITSTRTWALPTYNNHLKYQISSASTGASND 133  
DB 890 -----VDDGDNNTDSDSLSGRLSTSLRD-----DTGTSHE 922  
QY 134 NHYFGYSTFWGYFDENRFHCHFSPRDQRLINNNGWFRPKRLNFKLFNIQVKEVTTNDGV 193  
DB 923 RTYHKY-----WDHNTAQYDGLPDARF-----VREKETTTNTKV 956  
QY 194 TTIANNLSTVQVFSDEYQLPYVLGSAHQCLPPFPADVFMIPQYGYLTLNNGSQAVGR 253  
DB 957 SSGWREHYETE-YDDTE-----GASTTFLPMRTD-----WGQSSVSD-----NR 997  
QY 254 SSFYCLEYFPQSOMLRTGNMFTFSYTFEVPFH---SSYAHQSQSLRLMPLIDQYLYL 309  
DB 998 CTTYGRAVNTDNYDSTGAQ-RWTVLQDQKHYSGCSSIADSKQ-DGYASTLYD----- 1049  
QY 310 NRTQNGSQAONKDLFSRGSFAGMSVQPKNWLPGPCYRQQRVSKTKTDNNNSNFTWTGA 369  
DB 1050 ----NATSVANKPV---DGNPT-----ESRTYTKSGSYRSTWSG- 1082  
QY 370 SKYNLNGRESIINPGTAMASHKDEDEKF-----FPMGVMIFGKESAGA--SNTALDNVM 422  
DB 1083 --YDAGRWVWSEBDG---SHNRTLTKKSPANTWPLNGVITVTPDPDGLPAHTAL----- 1132  
QY 423 ITDEEEIKATNPVATERFGTVAVNFQSSSTD PATGVDVHAMGALPGVWQ 471  
DB 1133 -----TSTAMTSRFWGKPTSIQDANGVNTKMSLDAAGRLV-EVWR 1171

## RESULT 12

US-10-062-254-254  
; Sequence 254, Application US/10062254  
; Publication No. US20020138882A1  
; GENERAL INFORMATION:  
; APPLICANT: Cahoon, Edgar B  
; APPLICANT: Falco, Saverio Carl  
; APPLICANT: Fang, Yiwen  
; APPLICANT: Hantke, Sabine S.  
; APPLICANT: Lee, Jian-Ming  
; APPLICANT: Li, Zhongsen  
; APPLICANT: Miao, Guo-Hua  
; APPLICANT: Morgante, Michele  
; APPLICANT: Niu, Xiping  
; APPLICANT: Odell, Joan  
; APPLICANT: Rafalski, Antoni  
; APPLICANT: Sakai, Hajime  
; APPLICANT: Zheng, Peizhong  
; APPLICANT: Zhu, Qun  
; TITLE OF INVENTION: Polynucleotides Encoding Proteins Involved in Plant Metabolism  
; FILE REFERENCE:  
; CURRENT APPLICATION NUMBER: US/10/062,254  
; PRIOR FILING DATE: 2002-02-01  
; PRIOR APPLICATION NUMBER: 09/630,346  
; PRIOR FILING DATE: 2000-07-28  
; PRIOR APPLICATION NUMBER: 60/146511  
; PRIOR FILING DATE: 1999-07-30  
; PRIOR APPLICATION NUMBER: 60/156006  
; PRIOR FILING DATE: 1999-09-23  
; PRIOR APPLICATION NUMBER: 60/156899  
; PRIOR FILING DATE: 1999-09-30  
; PRIOR APPLICATION NUMBER: 60/157287

; PRIOR FILING DATE: 1999-10-01  
; PRIOR APPLICATION NUMBER: 60/169767  
; PRIOR FILING DATE: 1999-12-09  
; PRIOR APPLICATION NUMBER: 60/171054  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: 60/172958  
; PRIOR FILING DATE: 1999-12-21  
; PRIOR APPLICATION NUMBER: 60/171515  
; PRIOR FILING DATE: 1999-12-22  
; PRIOR APPLICATION NUMBER: 60/173535  
; PRIOR FILING DATE: 1999-12-29  
; NUMBER OF SEQ ID NOS: 375  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 254  
; LENGTH: 420  
; TYPE: PRT  
; ORGANISM: Zea mays  
US-10-062-254-254

Query Match 3.3%; Score 106; DB 15; Length 420;  
Best Local Similarity 20.4%; Pred. No. 0.78;  
Matches 76; Conservative 54; Mismatches 141; Indels 102; Gaps 19;

QY 44 SVDPDP-----LGEPPATPAVG--PTTMASGGGAPMADNNEGADVGNASGNWCH 93  
DB 54 AAPPPPPQFAKHFAAGPPPPQAGARRPTTPSSSGGNGCEENK----- 99  
QY 94 DSTWLGDRVITSTRTWALPTYNNHLKYQISSASTGASNDNHFGYSTPWGYFDENRFHC 153  
DB 100 -TIWGD-----LQYWMDENYLHNCFGPSGEVVTIKYIRNRQTQSGEGYGFVEF----- 147  
QY 154 HFSPRDWQRLINNNGF-----RPKRLNFKLFNIQVK--EVTINDG--VTTIANNLTS- 202  
DB 148 -FSHSAEKAQLQNTGVMPTDRPFKLNWASYSMGEKREYASDHSIFVGLADAVTDE 206  
QY 203 -TVQVFSDEYQLPYVLGSAHQCLPPFPADVFMIPQYGYLTL--NNGSQAVGR-SSFY 257  
DB 207 MLELFS-SKYR-----SVGAKAVIIDANTGRSRGYGFVRFGDDNDKSHAMSEMGVY 258  
QY 258 C---LEYFPQSOMLRTGNMFTFSYTFEVPFHSSYAHSQ-----SLDRLMN-- 299  
DB 259 CSTRPRLIGPATPRRSSGD-----SGSSTPGHSDGSSNRTVYVGLDPNVSEDELRYAF 313  
QY 300 -----PLIDQ--LYYLNRTQNGSQAONKDLFSRGSFAGMSVQPKNWLPGPCY 347  
DB 314 AKYDLASVKIPLGKQCGFVQFVSRDTAEBA-----LQGLNGSLIGQAVRLSWVRSPSH 367  
QY 348 RQGRVSKTKTDNN 360  
DB 368 KQSRGDSVNRNN 380

## RESULT 13

US-10-046-583A-1  
; Sequence 1, Application US/10046583A  
; Publication No. US20020168743A1  
; GENERAL INFORMATION:  
; APPLICANT: Rice, John  
; APPLICANT: Klotz, Andreas  
; APPLICANT: Crawford, John  
; APPLICANT: Lanning, Beth  
; APPLICANT: Stewart, Sandy  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE  
; TITLE OF INVENTION: IDENTIFICATION OF MODULATORS OF DEOXYXYLULOSE 5-PHOSPHATE  
; FILE REFERENCE: 2037 US Divisional  
; CURRENT APPLICATION NUMBER: US/10/046,583A  
; PRIOR FILING DATE: 2001-10-26  
; PRIOR APPLICATION NUMBER: 09/626,589  
; PRIOR FILING DATE: 2000-07-27  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1



LENGTH: 717  
TYPE: PRT  
ORGANISM: Arabidopsis thaliana  
US-10-046-583A-1

Query Match 3.2%; Score 104.5; DB 15; Length 717;  
Best Local Similarity 20.4%; Pred. No. 2.4;  
Matches 77; Conservative 54; Mismatches 121; Indels 125; Gaps 18;

QY 187 VTNDGVT--IANNLTSTVQVFSDEYQLPYLGSAGQCLPPFPADVEMIPQGYLT 243  
DB 12 ITTKGLSTDSCKSTSLSSSRLVTD---LP-----SPCLKP----- 45  
QY 244 LNNGSQAVGRSSFYCL-----EYF---PSQMLRTGNFTFSYTFEEVPEHSSYAHQS 293  
DB 46 -NNNSHS-NRRAKVCASLAEKEGEYYSNRPTPLDTIN-----YPIHMKNLVKE 93  
QY 294 LDRLMNPLIDQYLYLNRTQNSGSA-----QNKDLFSRGSFAGMSV 336  
DB 94 LKQSLDELRSQVTFNVSKTGHLGSSGLGVVELTVALHYIFNTPODKILM-----DVGHQS 148  
QY 337 QPKNWLPGPCYRQQRVSKTNDNNSNFTWTGASKYNLNG---RESTINPGTAMASHKD 393  
DB 149 YPHKILTG--RRGKMPTRQTNGLSGFTKRGESEHDCFGTGHSTTISAGLMAVGRDL 205  
QY 394 EDKFFPMGVMIFGKESAGASNTALDNVMTDEEIKATN-----PVAT-----ERFG 441  
DB 206 GKNNNVAVIGDAMTAGQAYEAMNAGYLDSDMIVILNDNKQVSLPTATLDGPPSPVG 265  
QY 442 TVAVNFQSSSTDPAT-----GDVHAMGA-----LPGMV-----WQDRD 474  
DB 266 ALSALSRLQSNPALRELRVAKGMTKQIGFPMQLAAKVDYARGMISGTGSSLFELG 325  
QY 475 VYLOGPIWAKIPHTDGH 491  
DB 326 LYYIGPV-----DGH 335

## RESULT 14

US-09-815-242-5845  
Sequence 5845, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
PRIOR FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 5845

LENGTH: 1016  
TYPE: PRT  
ORGANISM: Staphylococcus aureus  
US-09-815-242-5845

Query Match 3.2%; Score 103.5; DB 10; Length 1016;  
Best Local Similarity 21.5%; Pred. No. 4.9;  
Matches 122; Conservative 68; Mismatches 196; Indels 181; Gaps 33;

QY 77 NNEGADG-----VGNASGNHCDSTWLGDVVITSTRWALPT---YNNHLYKQISSAS 127  
DB 189 NTTGADGWGFLFSKGNAB-EYLTNGGILGDKLVNS-GGFKIDTGYITYTSSMDKTEKQAG 246  
QY 128 TGASNDNHYFGYSTPWGYFDENFRHCHFSPRDWQRLINNMGFRPKRLNFKLFNIQVKEV 187  
DB 247 QG-----YRGY---GAF-----VKNDSSG-----NSQWGENIDKS 274  
QY 188 TTNDGVTTIANNLSTVQVFSDEYQ-----LPYVLGSAGQGC----- 225  
DB 275 KTN--FLNYADNSTNT---SDGKHHGQRLNDVILTYVASTGKRAEYAGKTWETSITDL 328  
QY 226 -LPPFPADVEMIPQGYLTLLNNGSQAVGRSSFYCLEYFPQMLRT---GNNFTFS----- 276  
DB 329 GLSKQAYNFLTSSQRWGLNQGINANG-----WMRTDLKGSSEFTTPEAPK 375  
QY 277 -----YTFEEVPEHSSYAHQSQSLDRLMNPLIDQYLYLNRTQNSG-----SAQN-- 321  
DB 376 TITELKVEEIPFKK-----ERKFNPLAPGTEKVTYR-EGQKEKITTPTLKNPL 426  
QY 322 KDLFSRGSF-AGMSVQPKNWL-----GPCYRQQRVSKTNDNNSNFTWTGASKYN 373  
DB 427 TGVIIISGEPKEEITKDPINELTXYGPETIAPGRDEFDPKLPPT-----GEKEE 475  
QY 374 LINGRESTINP--GTAMASHKODEDKFPMGVMIFGKESAGASNTALDNVMTDEEIKATN 431  
DB 476 VPGKPGIKNPETGDVVRPVPDVSVTKYGPVKGDSIVEKEE-----IPFEKERKF 523  
QY 432 TNPVA--TERFGTVAVNFQSSST-----DPATGDVHAMGALPGMWQD--RDVYLOGPIW 482  
DB 524 NPDLAPGTEKVTREGQKEKITTPTLKNPLTGVIIISGSKSEKIEITKDPINELTXYGPET 583  
QY 483 AKIPHTDGHFP-----SPLMGFGFLKNPPQILIKNTIPVAPNPAEFSATKFAFIT 535  
DB 584 ITPGHRD-EFDPKLPTEKEKEVPKGPQIKNPETGDVVR-----PPVD-SVTKYGPV-- 632  
QY 536 QYSTQVSVIEIWELOKENSRRNDEV 562  
DB 633 ---KGDSIVEKE-EIPFEKERKFNPD 655

## RESULT 15

US-10-198-070-69  
Sequence 69, Application US/10198070  
Publication No. US20030109437A1  
GENERAL INFORMATION:  
APPLICANT: AVERBACK, PAUL  
APPLICANT: GEMMELL, JACK  
TITLE OF INVENTION: PEPTIDES EFFECTIVE IN THE TREATMENT OF TUMORS AND OTHER  
CONDITIONS REQUIRING THE REMOVAL OR DESTRUCTION OF  
FILE REFERENCE: 59003.000008  
CURRENT APPLICATION NUMBER: US/10/198,070  
PRIOR FILING DATE: 2002-07-19  
PRIOR APPLICATION NUMBER: 60/306,161  
PRIOR FILING DATE: 2001-07-19  
PRIOR APPLICATION NUMBER: 60/306,150  
PRIOR FILING DATE: 2001-07-19  
PRIOR APPLICATION NUMBER: 60/331,477  
PRIOR FILING DATE: 2001-11-16  
NUMBER OF SEQ ID NOS: 125  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 69  
LENGTH: 1265







GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 2003, 18:28:05 ; Search time 21.473 Seconds  
(without alignments)  
2681.720 Million cell updates/sec

Title: US-09-807-802A-15

Perfect score: 3251

Sequence: 1 TAPGKKRPVEQSPQEPDSSS.....NGLYTEPRPIGTRYLTRPL 599

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 73:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	ID	Description
1	2177.5	67.0	504	1 VCPV3A	coat protein - ade
2	1701	52.3	732	2 S52210	coat protein VP1 -
3	556	17.1	673	1 VCPVB5	coat protein VP1 -
4	495	15.2	781	1 VCPV19	coat protein VP1 -
5	258.5	8.0	723	1 VCPVFP	coat protein VP1 -
6	254.5	7.8	729	1 VCPVNA	coat protein VP1 -
7	254.5	7.5	722	1 VCPV2	coat protein VP1 -
8	244	7.2	587	1 B44276	coat protein VP1 -
9	234	6.6	716	1 VCPV2M	coat protein VP1 -
10	216	6.6	727	1 VCPV1F	coat protein VP1 -
11	214.5	6.6	718	1 VCPVIM	coat protein VP1 -
12	214	6.6	722	1 VCPVME	coat protein VP1 -
13	209	6.4	727	1 VCPVFP	coat protein VP1 -
14	208	6.4	737	1 VCPVCD	coat protein VP1 -
15	204	6.3	748	1 VCPVCP	coat protein VP1 -
16	198	6.1	584	2 S49594	capsid protein VP2
17	192	5.9	722	1 VCPVCN	coat protein VP1 -
18	191	5.9	702	1 VCPVAP	coat protein VP1 -
19	143.5	4.4	648	2 S50856	whn protein - rat
20	129	4.0	1072	2 A86827	hypothetical prote
21	123.5	3.8	690	2 AB0124	probable TonB-depe
22	123	3.8	1142	2 T37455	enamelin precursor
23	122.5	3.8	880	1 SYBSVS	valine-tRNA ligase
24	121	3.7	635	2 P96660	protein F2K1.10 f
25	121	3.7	1113	2 S28925	nuclear pore compl
26	120.5	3.7	931	2 T49710	related to glucan
27	119	3.7	667	2 A41311	transcription fact
28	118.5	3.6	1180	2 E86719	hypothetical prote

30	118	3.6	739	2 T52289	probable transket
31	116.5	3.6	642	1 S34416	transcription fact
32	116.5	3.6	1338	2 T30565	MAP kinase kinase
33	115.5	3.6	345	1 G97024	probable phosphoe
34	114	3.5	956	2 T08144	myosinase-binding
35	114	3.5	2271	2 P90073	hypothetical prote
36	113.5	3.5	2817	2 B97033	uncharacterized pr
37	113	3.5	1071	2 E85343	hypothetical prote
38	112.5	3.5	666	2 I52648	class A helix-loop
39	112.5	3.5	3078	2 T28432	variant-specific s
40	111	3.4	1777	2 T34369	hypothetical prote
41	110.5	3.4	655	1 ALKBG	cyclomaltodextrin
42	110.5	3.4	1742	2 T17120	cellulase (EC 3.2.
43	110	3.4	868	2 A82515	conserved hypothet
44	110	3.4	1296	2 C82521	hemolysin-type cal
45	110	3.4	2825	2 T14271	Doc4 protein, stre

ALIGNMENTS

RESULT 1									
VCPV3A									
coat protein - adeno-associated virus type 2									
C:Species: adeno-associated virus type 2									
C:Date: 05-Apr-1983 #sequence_rev1sion 05-Apr-1983 #text_change 16-Jul-1999									
C:Accession: A03698									
R:Srivastava, A.; Lusby, E.W.; Berns, K.I.									
J. Virol. 45, 555-564, 1983									
A:Title: Nucleotide sequence and organization of the adeno-associated virus 2 genome.									
A:Reference: A03694; MUID:83164299; PMID:6300419									
A:Status: translation not shown									
A:Molecule type: DNA									
A:Residues: 1-504 <SRI>									
A:Cross-references: EMBL:J01901; NID:G209616; PIDN:AAA42376.1; PID:G209621; EMBL:M12405;									
C:Superfamily: adeno-associated virus coat protein									
C:Keywords: coat protein									
Query Match									
Best Local Similarity 67.0%; Score 2177.5; DB 1; Length 504;									
Matches 392; Conservative 33; Mismatches 47; Indels 3; Gaps 2;									
QY	66	MASGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVITTSRTWALPTYNNHLYKQISS	125						
DB	1	MATGSGAPMADNNEGADGVGNAGNWHCDSTWMDGRVITTSRTWALPTYNNHLYKQISS	60						
QY	126	ASTGASNDNHFGYSTPMWGFDFNRFHCHFSPRDWORLNNWGFRLNFKLFNIQVK	185						
DB	61	QS-GASNDNHFGYSTPMWGFDFNRFHCHFSPRDWORLNNWGFRLNFKLFNIQVK	119						
QY	186	EVTNDGVTTIANNLTSTVQVFSDEYQLPYVLGSAHQGCLPPFPADVFMIPQGYLTIN	245						
DB	120	EVTQNDGTTTIANNLTSTVQVFTDSEYQLPYVLGSAHQGCLPPFPADVFMIPQGYLTIN	179						
QY	246	NGSQAVGRSSFYCLEYFPSSQMLRTGNNFTFSYTFEEVPPHSSYAHQSQSLDRLMPLIDQY	305						
DB	180	NGSQAVGRSSFYCLEYFPSSQMLRTGNNFTFSYTFEDVPPHSSYAHQSQSLDRLMPLIDQY	239						
QY	306	LYYLNFTQNGSGAQNKDLFSRGSFAGMSVQPKNMLPGPCYRQORVSKTKTDNNNSNFT	365						
DB	240	LYYLSRTNTPSGTTTQSRLOFSGAGASDIRDOSRNWLPGPCYRQORVSKTSADNNNSEYS	299						
QY	366	WTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMSGVMIFGKESAGASNTALDNVMTID	425						
DB	300	WTGATKHLNGRDSLVP--AMASHKDEDEKFFPMSGVILFGKQSEKTNVNIKVMITD	357						
QY	426	EEETKATNPVATERFGTVAVNFQSSSTDPAAGDVHAMGALPGWVWQDRDVLQGPIMAKI	485						
DB	358	EEETGTNPVATEQYGSVSTNLQRGNROATADVNTQGVLPQWVWQDRDVLQGPIMAKI	417						
QY	486	PHTDGHFHSPLMGFGGLKNPPQILLIKNTVPANPPAFSATKFAFITQYSTG	540						

Db 418 PHTDGHFHPSPLMGFGCLKHPPQILIKNTFPVPANPSTTFSAAKFAFITQYSTG 472

RESULT 2

S52210  
coat protein VP1 - muscovy duck parvovirus  
N/Alternate names: VP1 protein  
C/Species: muscovy duck parvovirus  
C/Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 17-Mar-2000  
C/Accession: S52210  
R/Zadori, Z.; Erdei, J.; Nagy, J.; Kisari, J.  
submitted to the EMBL Data Library, September 1993  
A/Reference number: S52209  
A/Accession: S52210  
A/Molecule type: DNA  
A/Residues: 1-732 <ZAD>  
A/Cross-references: EMBL:X75093; NID:g609091; PIDN:CAA52984.1; PID:g609093  
A/Experimental source: strain FM  
C/Genetics:  
A/Gene: VP1  
C:Superfamily: parvovirus coat protein  
C:Keywords: coat protein

Query Match 52.3%; Score 1701; DB 2; Length 732;  
Best Local Similarity 53.0%; Pred. No. 7,1e-109;  
Matches 325; Conservative 83; Mismatches 165; Indels 40; Gaps 11;

QY 6 KRPEQSPQEPDSSSGIGK-TGQOPAKKRLNFGQTGDSSEVPDPQLG-----E 53  
Db 141 EEPVNTAPAKKSS-----GKLTDDHPVKKPKLSE-----ENSPSPNSGGASAATEGSE 192  
QY 54 PPATPAVGPFTMASGGAPMADNNEGADGVGNAGNMHCDSTWLGDRVITTTSTRTWALP 113  
Db 193 PVAAP-----NMAEGSGAMGDSAGADGVGNAGNMHCDSQWLGDVYITTKTRTWLP 246  
QY 114 TYNNHLKYQISSASTGASNDNHFGYSTPWGYFDENRFCHFSFRDQRLINNMGFRPK 173  
Db 247 SYNMMYQAITSGTNPDSN-TQYAGYSTPWGYFDENRFCHFSFRDQRLINNMGIRPK 305  
QY 174 RLNFKLENIQVKEVTTNDGVTIANNLSTVQVFSDEYQLPYVLGSAHQCLPPFPADV 233  
Db 306 ALKFQIFNVQKEVTTQDOTKTIANNLSTIQIFTDNEHQLPYVLGSAHQCLPPFPADV 365  
QY 234 FMIPQGYLTN---NGSQAVGRSSFYCLEYFPSSQMLRTGNFTFSYTFEEVFPFHSYAH 290  
Db 366 YALPQGYCTWHTNQSARFNDRSAFYCLEYFPSSQMLRTGNFTFSYTFEEVFPFHSYAH 425  
QY 291 SOSLDRLMPLIDQYLYLNRTONQSGSAQNKDLFSRGSPPAGMSVQPKMWLPGPCYRQ 350  
Db 426 SQDLRLMPLIDQYLYLNFSEV-NGGRNAQ-----FKKAVKGAFGAMGRNWLPGPKLLDQ 479  
QY 351 RVSKTK--TDNNNSNFTWTGASKYNLNGRESINPGTAMASHKDDDEKFFPMGVMIFGK 408  
Db 480 RVRAYSGGTDNVANWSIMSKGNKFLKDRREYLLQPGVATHTTEDQASSVPAQNIIGIAK 539  
QY 409 E--SAGASNTALDNVMTDEEIKATNPVATERFCTVAVNFQSSSTDPAIGDVHAMGALP 466  
Db 540 DPYRSGSTLAGSDIMWTEQEIAPTNGVGRPFYGLTVTNEQNTTAPINAELEVLGALP 599  
QY 467 GMVWQDRVYLOGPIWAKIPHTDGHFHPSPLMGFGCLKHPPQILIKNTFPVPANPAPFES 526  
Db 600 GMVWQNRDIYQGIWAKIPKTDGKPHSPNLGGFGLHNPFOVFIKNTFPVPADPLEYV 659  
QY 527 ATKFASFIQYSTGVSVIEIEMELQKENSKRNPDEVQTSNYAKSANDVFTVDNNGLYTE 586  
Db 660 NQKWNSSYITQYSTGQCTVEMVWELRKENSKRNPDEIQTSNFGNRTSTMFAPNETGYVE 719  
QY 587 PRPIGTRVLTPL 599  
Db 720 DRLIGTRVLTQNL 732

RESULT 3

VCPVB5  
coat protein VP1 - bovine parvovirus  
N/Contains: coat protein VP2  
C/Species: bovine parvovirus  
C/Date: 28-Dec-1987 #sequence\_revision 28-Dec-1987 #text\_change 16-Jul-1999  
C/Accession: A26104  
R/Chen, K.C.; Shull, B.C.; Moses, E.A.; Lederman, M.; Stout, E.R.; Bates, R.C.  
J. Virol. 60, 1085-1097, 1986  
A/Title: Complete nucleotide sequence and genome organization of bovine parvovirus.  
A/Reference number: A26104; MUID:87061184; PMID:3783814  
A/Accession: A26104  
A/Molecule type: DNA  
A/Residues: 1-673 <CHB>  
A/Cross-references: EMBL:M14363; NID:g333454; PIDN:AAB59847.1; PID:g808805  
C:Superfamily: parvovirus coat protein  
C:Keywords: coat protein  
F/138-673/Product: coat protein VP2 #status predicted <VP2>

Query Match 17.1%; Score 556; DB 1; Length 673;  
Best Local Similarity 25.0%; Pred. No. 2.6e-30;  
Matches 160; Conservative 98; Mismatches 239; Indels 144; Gaps 19;

QY 1 TAPGKKRPEQSPQEPDSSSGIGKTGQOPAKKRLNFGQTGDSSEVPDPQLGEP--PATP 58  
Db 100 TSKGGBRALKRLYFARSNKAKKANREPASTSNQONNEVNDIPNDEAGNQIELATR 159  
QY 59 AAVGPTMASGGGAPMADNNEGADGVGNAGNMHCDSTWLGDRVITTTSTRTWALPTVNNH 118  
Db 160 SVVGGSGVGGG-----RGSGGVYSTGWTGTTFSENIVTKNTRQFICIDIKNGH 211  
QY 119 LYKQISSASTGASNDNHFGYSTPWGYFDENRFCHFSFRDQRLINNMGFRPKRLNFK 178  
Db 212 LYKS-EVLNTGDTAHRQY-AITFPWSYENFNQYSSHSFPNDQHLVNDYERFRPKAMAYR 269  
QY 179 LFNIDYKEVTTNDGVTIANN-LTSTVQVFSDEYQLPYVLGSAHQCLPPFPADVEMIP 237  
Db 270 VYNLQIKQIMTDGAMGTVYNNDLTAGMHIFCDGDHRYPYQHPPWDQCMPELPNSIWELP 329  
QY 238 QYGYLT-----NGSQAVGRS-----SFYCLEYFPSSQMLRTGNFTFSYTFEEVFPFHS 286  
Db 330 QYAYIPAPISVDNNTTNTVEEHLKGVPLMYLLENSDHEVLKNG----- 373  
QY 287 SYAHSQSLRLMPLIDQYLYLNRTON-----QSGSAQNKDLFSRGS----- 331  
Db 374 -----RIVRIYQLWRLMDRKQHHIQHASDVQSTQKQKNLLIQRTKQPNKQRF 424  
QY 332 AGMSVQPKMWLPGPCYRQQRVSKTKTDNNNSNFTWTGASKYNLNGRESIINPGTAMASHK 391  
Db 425 QNAALRTSNMMSGP-----GIARGTHNATLQTSAGALVTWVT 462  
QY 392 DDEKFFPMGVM-----IFGKE-----SAGASNTALDNVMTDE 426  
Db 463 NGAD-----VSGVRAVRVGYSTPIYGGQQPESDLRLRYSASAAEQQNPILEN----- 512  
QY 427 EEIKATNPVATERFCTVAVNFQSSSTDPAIGDVHAMGALPGMWQDRDVLQGIWAKIP 486  
Db 513 -----AARHTFTREARTIKLITGSGNADGDYKEWMLPNQWDSAPISRNPWIKVP 564  
QY 487 HTDGHFHPSPLMGFGCLKHPPQILIK--NTPVPANPAPFESATKFAFITQYSTGVSV 544  
Db 565 RVNRKTLTLDQDSIPMSHPGTFIKLARIPVPGND-----SFLNIYVTGQVSC 615  
QY 545 EIEIEMELQKENSKRNPDEVQTSNYAKSANDVFTVDNNGLY 584  
Db 616 EVVMEVEKRGTKNMRPEYMHs---ATNMSVDAYTINNAGVY 653

RESULT 4

VCPV19  
coat protein VP1 - parvovirus B19 (strain Au)  
C/Species: parvovirus B19  
A/Note: host Homo sapiens (man)  
C/Date: 28-Dec-1987 #sequence\_revision 28-Dec-1987 #text\_change 16-Jul-1999









Query Match 6.6%; Score 216; DB 1; Length 716;  
Best Local Similarity 20.8%; Pred. No. 6.3e-07;  
Matches 148; Conservative 93; Mismatches 271; Indels 198; Gaps 34;

```
QY      8 PVEQSPQEPDSSSGIGKTG-----QQPAKKRL-----NFGQT-GDSESVDP 48
      82 PKLATDSEF-GTSGVSRAGKRTTPPAYIFINQARAKKULTSSAQSSQTMDSGTSDPS 140
QY      49 QPLGEPRA--TPRAVGPFTMASGGGAPMADNNEGADGVNASGNMHCDS--TWLGD--R 101
      141 GNAVHSAARVERAADPG--GSGGG-----GSGGGGVGVSTGSDYDQTHYRFLGDGWVE 192
QY      102 VTTSTRTWALPTYNHLYKQI-----SSASTGASNDNHYPGYSTPWGYFDENRFCH 154
      193 ITALATRLVHLNMPKSENYCRIRVHNTTDSYKGNMAKDDAHQIWTWPSLVDANAWGVW 252
QY      155 FSPRDMQRLINNMGFRPKRLNFKLFNIQVKEVTND---GVTTIANNLSTVQVFSDS 210
      253 LQPSDWQYICNTMSQNLVSLDQEIFNVVLKVTEDLGGQAIKIYNNDLTACMMVAVDS 312
QY      211 EYQLPYVLGSAHQCLPFPADVEMIPQGY-----LTLNNGSQAV-----G 252
      313 NNILPYTPAANSMETLGFYPMKPTIASPRYRYFCVDRDLSTVENQEGVEHNVMTGTPKG 372
QY      253 RSSFYCLEYFPS-QMLRTGNMF-TFSYTFEEVFPFHSYAHQS3LDRLN--PLIDQYLYY 308
      373 IPOFTIENTQQTILRLTGTDEPATGTYYPFDINSV--KLTHWTQTNRLQGPPLLSTF--- 427
QY      309 LNRTONGSQAQNDLFSRGPAGMSVQPKMWL-----PG 344
      428 -PEADTDAGT-----LTAQGSRHGTTQMGVNVWVSEAIRTPAQVGFQCPHNDFEASRAG 480
QY      345 PCYRQORVSKTKTDNNNSNFTWTGASKYNLNGRESIINPG-----TAMASHKDE 394
      481 P-FAAPKVPADITQGVNKEANGSVRYSGKQHGEMWASHGPAPERYTWDETSFGSGRDTK 539
QY      395 DKFF-----PMSGVMIFGKESAGASNTALDNWMITDEEIKATNPVATERFGTVAV 445
      540 DGFISQAPLVPPLNGI-----LTNANPIGTKN---DI 570
QY      446 NFQS--SSTDPAFGDVHAMGALPGMWQDRDVLQGFPIWAK--IPHTDGHFHPSPLMGG 500
      571 HFSNVFNSYGPLTAFSH-----PSPYVPOQOIWDKELDLHKRRLHITAPFV-- 617
QY      501 FGLKNPPQILIKNTP--VPANPPAEFSATKPAFITQYSTGQVSEIEMELQKENSKRW 558
      618 -CKNNAPGQMLVRLGPNLTDQYDPNGATLSRIYTYGTFPWKGLTMRAKLRA---NTTW 672
QY      559 NPEVQYTSNYAKSANVDFTVDNNGLY---TEPRPIGT-----RYLTRPL 599
      673 NPVYQVSAAE-----DNGNSYMSVTWKWLPATGNMQSVPLITRPV 711
Db
```

RESULT 11  
VCPV1P  
coat protein VP1 - feline panleukopenia virus

N/Contains: coat protein VP2  
C/Species: feline panleukopenia virus, FPLV  
C/Date: 30-Jun-1987 #sequence\_revision 30-Jun-1987 #text\_change 16-Jul-1999  
C/Accession: A03701  
R/Carlson, J.; Rushlow, K.; Maxwell, I.; Maxwell, F.; Winston, S.; Hahn, W.  
J. Virol. 55, 574-587, 1985  
A/Title: Cloning and sequence of DNA encoding structural proteins of the autonomous parvovirus  
A/Reference number: A03697; MUID:85265017; PMID:2991581  
A/Accession: A03701  
A/Molecule type: DNA  
A/Residues: 1-727 <CAR>  
A/Cross-references: EMBL:M10824; NID:g333474; PIDN:AAA47161.1; PID:g333476  
C/Genetics:  
A/Intons: 11/1  
C/Superfamily: parvovirus coat protein  
C/Keywords: coat protein

F,144-727/Product: coat protein VP2 #status predicted <VP2>

Query Match 6.6%; Score 216; DB 1; Length 727;  
Best Local Similarity 20.2%; Pred. No. 6.4e-07;  
Matches 139; Conservative 101; Mismatches 259; Indels 188; Gaps 34;

```
QY      7 RPVEGSPQEPDSSSGIGKTGQQPAKKRLNFGQTDSESV----DQPLGEPRA--TPAA 60
      110 KPTKRSKPPPHIFINLAK-----KKKAGAQVKRDNOAPMSDGAQVDPDGGQPAVNERA 163
QY      61 VGPTTMASGGGAPMADNNEGADGVNASGNMHCDS--TWLGD--R 113
      164 TGSNGSGGGG-----GSGSGGISTGTENNQTEFKFLENGWV--EITANSSRLVHLN 215
QY      114 TYNNHLYKQI-----SSASTGASNDNHYPGYSTPWGYFDENRFCHFSPRDMQRLINN 166
      216 MPSENYKRVVNNMDKTAVKGNMALLDTHVQIVTPWPSLVDANAWGVWFPNGDWQLIVNT 275
QY      167 NMGFRKRLNFKLFNIQVKEVT--TNDGVTTIANNLSTVQVFSDEYQLPYVLGSAHQ 223
      276 MSELHVSFEQEIFNVVLKTVSESATQPTPKVYNNDLTASLVALDSNNTMFTPRAMRS 335
QY      224 GCL-----PPFPA-----DVEMIPQGYLTLNNGSQAVGRS-----SFYCL 259
      336 ETLGFYPMKPTIPTPMRYFPQMDRTLIPSH-----TGTSQPTNIIYHGTDPDDVQFYTI 389
QY      260 E-YFSPQMLRTGNMF-TFSYTFEEVFPFHSYAHQS3LDRLN--PLIDQYLYLNRTQNG 315
      390 ENSVPVHLRLTGTDEPATGTFPFDCKP--CRLTHWTQTNRALGLP-----FLNSLPQS 440
QY      316 SGSQAQNDLFSRGPAGMS-----VQPKN-WLPGPCYRQORVSK----- 354
      441 EGATNFGDIGVQDQKRGVTOGNTDYITEATIMRPAEVGYSAFYSEASTQGFKPI 500
QY      355 -----TKTDNNNSNFTWTGASKY--NLNGRESIINPGT---AMASHKDEDEKFFPMS 401
      501 AAGRGACTDENQAA---DGDRYAFGROHQKTTTGTGTPERTYIAHODT----- 549
QY      402 GVMIFKESAGASNTALD-NWMITDEEIKATNPVATERFGTVAVNFQSSSTDPAFGDVH 460
      550 -----GRYPAGDWIQININFLPVTDNVLPLTDPIG---GKTGINY--TNIFNTYGPILT 598
QY      461 AMGALPGMWQDRDVLQGFPIWAKIPHTDGHFHPSPLMGGGLK-----NPP 507
      599 ALNNVP-----PVYPNGQIWDKEFDTD-----LKPRLHVNAPFVCQNNCP 638
QY      508 PQLIKNTPVPAN--PPAEFSATKPAFITQYSTGQVSEIEMELQKENSKRWPEVQY 564
      639 GQLFVYKAPNLITNEYDPDASANSR---IVTYSDFWWMKGLVFKAKL.RASHWTWNP1QOM 694
QY      565 TSNYAKSANVDFTVDNNGLYTEPRPIG 591
      695 SIN-----VDNQFNVL-PNNIG 710
Db
```

RESULT 12  
VCPV2M

coat protein VP1 - minute virus of mice (strain MVM1)

C/Species: minute virus of mice, murine parvovirus  
C/Date: 28-Dec-1987 #sequence\_revision 28-Dec-1987 #text\_change 28-Jul-2000  
C/Accession: B23008; B29510  
R/Sahli, R.; McMaster, G.K.; Hirt, B.  
Nucleic Acids Res. 13, 3617-3633, 1985  
A/Title: DNA sequence comparison between two tissue-specific variants of the autonomous parvovirus  
A/Reference number: A23008; MUID:85242059; PMID:3855242  
A/Accession: B23008  
A/Molecule type: DNA  
A/Residues: 1-718 <SAH>  
A/Cross-references: EMBL:X02481; NID:g60918; PIDN:CAB46507.1; PID:g5419928  
R/Asfari, C.R.; Gardner, E.M.; Tattersall, P.  
J. Virol. 57, 656-669, 1986  
A/Title: DNA sequence of the lymphotropic variant of minute virus of mice, MVM(i), and cc  
A/Reference number: A29510; MUID:86115415; PMID:3502703



A:Accession: B29510  
A:Molecule type: DNA  
A:Residues: 1-143, 'A', 145-718 <AST>  
A:Cross-references: EMBL:M12032  
C:Superfamily: parvovirus coat protein  
C:Keywords: coat protein

Query Match 6.6%; Score 214.5; DB 1; Length 718;  
Best Local Similarity 21.0%; Pred. No. 8e-07;  
Matches 142; Conservative 89; Mismatches 273; Indels 171; Gaps 30;

```
QY 8 PVEQSPQEPDSSSGIGKTG-----QOPAKRLNFG---QTGDSSEVPDPQLG 52
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 83 PKLATDSEP-GTSGVSRAGKRTTRPAYIFINQARAKKLTSSAQSSQTMDSQSDG 141
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 53 -----EPPATPAVGPPTMASGGAPMADNNEGADGVGNASGNWHGDS--TWLGD---R 101
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 142 GNGVHSARVERADPG--GSGGG-----GSGGGGVGVSSTGSDNQTHYRFLGDGWE 193
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 102 VITSTRTVALPTYNHLYKQI-----SSASTGASNDNHFGYSTPWGYFDENRFCH 154
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 194 ITLATATRLVHLMPSKSENYCRIRVHNTTDSVKGNMAKDDAHEQIWTWPSLVDANAMGW 253
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 155 FSPRDWQRLNNWGFPRKRLNFKLENIQVKEVTND---GVTTIANNLSTVQVFSDS 210
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 254 LQPSDWQICNTMSQLNLVSLDQEIFNVVLKTVTEQDSGGQAIKIYNNDLTACMMVAVDS 313
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 211 EYQLPYVLGSAHQCLPPFPADVFMIPQGY-----LTINN-----GSAV 251
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 314 NNILPYTPAANSMETLGFYPMKPTIASPRYYFCVDRDLSTYENQEGTIEHNVMTGPKG 373
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 252 GRSSFYCLEYFPS-QMLRTGNNE-TFSYTFEEVFPFHSYAHQSGLDRLMN--PLIDQYLY 307
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 374 MNSQFFLENTQQITLLRTGDEFATGYFEDTNPV--KLTHTWQTNRLQGPPLSTF-- 429
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 308 YLNRTQNGSQAQNKDLLFSRGSPPAGMSVQPKNL-----P 343
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 430 --PEADTDACT-----LTAQGRHGATQMEVNVWVSEAIRTPRAOVGFCQPHNDFEASRA 481
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 344 GPCYRQQRVSKTKTDNNSNFTWTGASKYNLNGRESIINPGTAMASHKDEDKF----- 397
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 482 GP-PAAPKYPADVTQGVDRANGSVRSYKQKHGENWAAHGAPAPERYTDETFNGSGRDT 540
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 398 ---PPMSGVMIFGKESAGASNTALDNVMTDEEIKATNPVATERFETVAVNFQS--SST 452
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 541 RDGFIQSAPLVPPPLNGILTNA-----NPIGTKN----DIHFSNVFNSY 581
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 453 DPATGVHANGALPGMWQDRDVLQGPIMAK--IPHTDGHFHSPLMGFGGLKXNPPQ 509
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 582 GPLTAFSH-----PSPVYPQGLMDKELDLHKPRLHITAPFV---CKNNAPGQ 627
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 510 ILIKNTPP--VPANPPAEFSATKFAFITQYSTGVSEIEMELQKENSKRWNPEVQYTSN 567
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 628 MLVRLGPNLTDQYDPNGATLSRIIVTYGTFFWKGLTMRAKLRA-----NTTNPPVYQ---- 679
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 568 YAKSANVDFTVDNNG 582
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 680 -----VSVEDNG 686
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

## RESULT 13

VCPVME

coat protein VP1 - mink enteritis virus (strain Abashiri)

N:Contains: coat protein VP2

C:Species: mink enteritis virus, MEV

C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 12-Apr-1996

C:Accession: B38350

R:Kariatsumari, T.; Horiuchi, M.; Hama, E.; Yaguchi, K.; Ishiguro, N.; Goto, H.; Shinag

J. Gen. Virol. 72, 867-875, 1991

A:Title: Construction and nucleotide sequence analysis of an infectious DNA clone of the

A:Reference number: A38350; MUID:91202123; PMID:2016597

A:Accession: B38350

A:Molecule type: DNA

A:Residues: 1-722 <KAR>  
A:Cross-references: GB:D00765  
C:Superfamily: parvovirus coat protein  
C:Keywords: coat protein  
F:139-722/Product: coat protein VP2 #status predicted <VP2>

Query Match 6.6%; Score 214; DB 1; Length 722;  
Best Local Similarity 20.2%; Pred. No. 8.7e-07;  
Matches 139; Conservative 101; Mismatches 259; Indels 188; Gaps 34;

```
QY 7 RPEQSPQEPDSSSGIGKTGOOPAKRLNFGQTGDSSEVP---DPQLGEPPA--TPAA 60
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 105 KPTKRSKPPPHIFINLAK-----KKKAGAGQVKRDNLA PMSDGAVQPDGGQPAVNERA 158
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 61 VGPTTASGGAPMADNNEGADGVGNASGNWH-----CDSTWLGDRTVITSTRTALP 113
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 159 TSGNGSGGGG-----GGSGGGVIGISTGFNNQTEPKFLENGWV--EITANSSRLVHLN 210
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 114 TYNNHLYKQI-----SSASTGASNDNHFGYSTPWGYFDENRFCHFSPRDWRILNN 166
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 211 MPSENYKRNVVNNMDKTAVKGNMALDDTHVQIVTWPMSLVDANAMGWFMNPGDWQLIVNT 270
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 167 NMGFRPKRLNFKLENIQVKEVT---TNDGVTIANNLSTVQVFSDEYQLPYVLGSAHQ 223
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 271 MSELHLVSFEQEIFNVVLKTVSESATQPTKVVYNNDLTASLWALDSNNTMPTPAMRS 330
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 224 GCL-----PPFA-----DVEMIPQGYLTLNNGSQAVGRS-----SFYCL 259
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 331 ETLGFYPMKPTIPTPMRYFQMDRTLIPSH-----TGTSPTTNHYGTDPPDVQFYTI 384
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 260 E-YFPSQMLRTGNNE-TFSYTFEEVFPFHSYAHQSGLDRLMN--PLIDQYLYLNRTQNO 315
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 385 ENSVAVHLRLTGDEFATGYFEDTNPV--CRLTHTWQTNRALGLP-----PLNSLPQS 435
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 316 SGSQONKDLLFSRGSPPAGMS-----VQPKN-WLPGCYRQQRVSK----- 354
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 436 EGATNFGDIGVQQDKRGVQNGNTDYITEATIMRPAEYGSAPYSSFEASTQGPFTPI 495
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 355 -----TKTDNNSNFTWTGASKY--NLNGRESIINPGT---AMASHKDEDKFPMBS 401
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 496 AAGRGCAQTDENQAA--DGDPRYAFGRQHGQKTTTGTETPERFTYIAHQDT----- 544
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 402 GVMIFGKESAGASNTALD-NVMTDEEIKATNPVATERFETVAVNFQSSTDPATGVH 460
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 545 -----GRYPAGDWIQINENLPTNDNLPLTDPDG-----GKTGINY--TNIFNTYGPLT 593
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 461 AMGALPGMWQDRDVLQGPIMAKIPHTDGHFHSPLMGFGGLK-----NPP 507
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 594 ALANNVP-----PYYPNGQIMDKFEDTD-----LKPRLHVNAFPVCQNNCP 633
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 508 PQILIKNTPPVAN--PPAEFSATKFAFITQYSTGVSEIEMELQKENSKRWNPEVQY 564
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 634 GOLFWKVAPNLTNEYDPDASANSR---IVTYSDFWKGKLVFKAKLRASHWTWNPLOQM 689
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 565 TSNYAKSANVDFTVDNNGLYTEPRPIG 591
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 690 SIN-----VDNQFNYL-PNNIG 705
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

## RESULT 14

VCPVFP

coat protein VP1 - feline panleukopenia virus (strain 193)

N:Contains: coat protein VP2

C:Species: feline panleukopenia virus, FPLV

C:Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 12-Apr-1996

C:Accession: B36608

R:Martyn, J.C.; Davidson, B.E.; Studdert, M.J.

J. Gen. Virol. 71, 2747-2753, 1990

A:Title: Nucleotide sequence of feline panleukopenia virus: comparison with canine parvov

A:Reference number: A36608; MUID:91073139; PMID:2174965

A:Accession: B36608

A:Molecule type: DNA

A:Residues: 1-727 &lt;MAR&gt;

A;Cross-references: GB:X55115  
C;Superfamily: parvovirus coat protein  
C;Keywords: coat protein  
F;144-727/Product: coat protein VP2 #s

Query Match 6.48; Score 209; DB 1; Length 727;

Best Local Similarity 20.1%; Pred. No. 1.9e-06;

Matches 138; Conservative 101; Mismatches 260; Indels 188; Gaps 34;

```

QY 7 RPEQSGPOEPDSSSGIGKTGQOPAKKRLNFGQTGDSSEVP---DPÖPLGEPPA--TPAA 60
Db 110 KPTKRSKPPPHIFINLAK-----KKKAGAGQYKRDNLAPMSDGAVÖPDDGÖQPAVNERA 163
QY 61 VGPTTASGGGAPMADNNEGADVGNASGNWH-----CDSTWLGDRIYTTSTRTWALP 113
Db 164 TGSNGSGGGG-----GGSGGVGISTGTENNÖTEPKFLENGWV--EITANSSRLVHLN 215
QY 114 TYNNHLKYÖI-----SSASTGASNDNHYPGYSTPMGYFDNRFCHFSPRDÖRLINN 166
Db 216 MPESENTKRVVNNMDKTAVKGNMALDDIHQIVTPWSLVDANAMGWENFPGMÖLIVNT 275
QY 167 NMGFRPKRLNFKLFNIÖVKEVT---TNDGVTTIANNLSTVÖVFSDEYÖLPVYLGAHQ 223
Db 276 MSELHLVSFEÖEIFNVVLKTVSESATÖPPTKYNNDLTASLMAVALDSNNTMPTIPAAMS 335
QY 224 GCL-----PPPPA-----DVFMIPÖGYLTLNNGSQAVGRS-----SFYCL 259
Db 336 ETLGFPYWKPTIPTPWRYFYÖWDRTLIPSH-----TGTSGPTNVHGTDPDDVÖFYTI 389
QY 260 E-YEPSÖMLRTGNNE-TFSYTFEEVPPHSSYAHSQSLDRLMN--PLIDÖLYLYNRTÖNÖ 315
Db 390 ENSVPEHLRTGDEFATGTFFDCKP--CRLTHTWÖTRALGLP-----FLNSLPÖS 440
QY 316 SGSAÖNKDLLFSRGSPPAGMS-----VÖPKN-WLPGPCYRÖÖRVSK----- 354
Db 441 EGATNPGDIGVÖQÖDKRGVTÖMGNTDYITREATIMRPAEVGSAPYYSFEASTGÖPEKPTPI 500
QY 355 -----TKTDNNNSNFTWTGASKY--NLNGRESIINPGT---AMASHKDEDEKFFPMS 401
Db 501 AAGRGGAÖTDENÖAA---DGDPRYAFGRÖHQÖKTTTGTETPERFTYIAHQDT----- 549
QY 402 GVMIFGESAGASNTALD-NVMITDEEBEIKATNPVATERFGTVAVNFÖSSSTDPATGDVH 460
Db 550 -----GRYPEGDWICÖNINFNLPVTDNVLLPTDPIG---GKTGINY--TNLFNTYGPULT 598
QY 461 AMGALPGMWVÖDRDVLÖGPIWAKIPHTDGHFHPSPLMGGFGLK-----NPP 507
Db 599 ALNNVDP-----PVYPNGÖIMDKEPDFTD-----LKPRLHVNAPEVCÖNNCP 638
QY 508 PÖLILKNTVPVAN---PPAEFSATKFPASFIÖYSTGÖVSEJEMELÖKENSKEWNPEVÖY 564
Db 639 GÖLPEVKVAPNLITNEYDPDASANMSR---IVTYSDFWWMKGKLVFKAKLRASHITNPIÖQM 694
QY 565 TSNYAKSANDFTVDNNGLYTEPRPIG 591
Db 695 SIN-----VDNÖFNTYV-PNNIG 710

```

## RESULT 15

VCPRVCD

coat protein VP1 - canine parvovirus (strain CPV-d)

N;Contains: coat protein VP2

C;Species: canine parvovirus, CPV

C;Date: 31-Mar-1990 #sequence revision 31-Mar-1990 #text: change 16-Jul-1999

C;Accession: A31163

R; Parrish, C.R.; Aquadro, C.F.; Carmichael, L.E.

Virology 166, 293-307, 1988

A; Title: Canine host range and a specific epitope map along with variant sequences in the

A:Reference number: A31163; MUID:89020796; PMID:3176341

A;Accession: A31163

A;Molecule type: DNA

A;Residues: 1-737 &lt;PAR&gt;

A;Cross-references: EMBL:M23255; NID:g3333467; PIDN:AAA47158.1; PID:g3333468

C;Genetics:  
A;Introns: 26/3  
C;Superfamily: parvovirus coat protein  
C;Keywords: coat protein  
F;584-737/Product: coat protein VP2 #status predicted <VP2>

Query Match	6.4%	Score 208	DB 1	Length 737
-------------	------	-----------	------	------------

Best Local Similarity 19.8%; Pred. No. 2.3e-06;  
Matches 137, Correctations 85, Mismatches 360

Matches 137; Conservative 96; Mismatches 260; Indels 200; Gaps 31;

QY	7	RPVEQSPOEPDSSSGIGKTGQOPAKKRLLNFQGTGDSSEVP----	DPQPLGEPPA--	TPAA	60		
Db	120	KPTKRSKPPPHIFINLAK-----	KKKAGAGQVKRDNLAPEMSDGA	VQPDGGQPAVRNERA	173		
QY	61	VGPPTMASGGGAPMADNNEGADVGNASGNWH-----	CDSTWLGD	RVITSTRTWALP	113		
Db	174	TGSGNGSGGG-----	GGGGGCVGISTGTENNQTEFKFLENGWV--	EITANSSRLVHLN	225		
QY	114	TYNNHLKQI-----	SSASTGASNDNHYFGYSTPWGYFDENRFCHFS	PRDWQRLINN	166		
Db	226	MESENRYRRVVNNMDKTA	VNGNMALDDIHAQIVT	PMSLV	DANAMGWENPGDWQLIVNT	285	
QY	167	NMGFRPKRLNFKLFENIQVEVT--	TNDGVTTIANNLTSTVOYFSDSEYQLPYVLGSAHQ		223		
Db	286	MSLHLVSFEQEIFNVVLKTVSESATQPTPKVYNNDLTASLMA	VALDSNNTMPTPAAMRS		345		
QY	224	GCL-----	PPPA-----	DVEMIPOCYLLTNNGSAVGRS-----	SFYCL	259	
Db	346	ETLGFYPMKPTIPTPMWRYXEQMDRTLIPSH-----	TGSGTPTNIYHGTDPDDVQFYTI		399		
QY	260	E-YEPSQMLRTGNNF-TFSYTFEEVPEFHSSYAH	SQSLDRLMN--	PLIDQLYYLNRTONQ	315		
Db	400	ENSVPVHLRLRTGDEFATGTFEEDCKP--	CRLTHTWQTRALGLP	-----	FLNSLPQS	450	
QY	316	SGSAQNKDLLFSRGS	PAGMSVQPKNWLPGPCYRQORVSKTKTDNNSNF-----	TW	366		
Db	451	EGATNFGDI-----	GV-----	QODKRRGVTQMGNNTNYITEATIMRPAE	488		
QY	367	TGASKYNLNGRESTIINP----	GTAMASHKDEDEKFFPMG	SVIMFGKESAGASNTALD--	419		
Db	489	VGYSA	PYYSFEASTQGP	FKTPIAAGRGGAQT	DENQALADGNPRYAFGRQHGOKTTTGETP	548	
QY	420	-----	NMITDEEIKATNPVATERFGT	VA	VNFQSSSTDP	454	
Db	549	ERFTYIAHQDTGRYPEGDWIONINENL	PVTNDVLLPTDP	IG----	GKTGINY--TNIFN	602	
QY	445	ATGDVHAMGALPGMWQDBDYVLQGP	IWAKIPHTDGHFHSPLMG	GGFLK-----		504	
Db	603	TYGPLTALNNVP-----	PYPNGQIMDK	EFTD-----	LKPRLHVNAPEV	642	
QY	505	---NPPQILIKNT	PVPAN----	PPAEFSATKFASFITQYSTGQV	SVEIEMELQKENS	KRW	558
Db	643	CONNCPGQLFVKVAPNL	THEYDPDASANMSR----	IYYSDFW	MWKGKLVFKAKLRASHTW	698	
QY	559	NBEVQYTSNYAKS	ANVDFTVDNNGLYTEPR	PIG	591		
Db	699	NPIQQMSIN-----	VDNQFN	YV--PSNIG	720		

Search completed: July 17, 2003, 18:35:46

```

          comp20000: 00473
Job time : 23.473 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 2003, 18:20:49 ; Search time 11.8582 Seconds  
(without alignments)  
2095.115 Million cell updates/sec

Title: US-09-807-802A-15  
Perfect score: 3251  
Sequence: 1 TAPGKKRPVEQSPQEPDSSS.....NNGLYTEPRPIGTIRYLTRPL 599

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2177.5	67.0	504	1 COA3_AAV2	P03135 adeno-asso
2	556	17.1	673	1 COAT_PAVBO	P07297 bovine parv
3	495	15.2	781	1 COAT_PAVHB	P07299 human parvo
4	259.5	8.0	749	1 COA2_PAVPN	P22964 porcine par
5	254.5	7.8	729	1 COA1_PAVPN	P33484 porcine par
6	254.5	7.8	729	1 COA1_PAVPN	P18546 porcine par
7	252.5	7.8	729	1 COA1_PAVPN	P52501 porcine par
8	244	7.5	722	1 COAT_PAVHH	P03136 hamster par
9	234	7.2	587	1 COAT_PAVL3	P36310 parvovirus
10	216	6.6	716	1 COAT_PAVL3	P03137 murine minv
11	216	6.6	727	1 COAT_PAVL3	P04864 feline minv
12	214.5	6.6	718	1 COAT_PAVL3	P07302 murine minv
13	214	6.6	722	1 COAT_PAVL3	P27437 mink enteri
14	209	6.4	727	1 COAT_PAVL3	P24840 feline parv
15	208	6.4	727	1 COAT_PAVL3	P17455 canine parv
16	208	6.4	737	1 COAT_PAVL3	P12930 canine parv
17	204	6.3	748	1 COAT_PAVL3	P30129 canine parv
18	195	6.0	584	1 COAT_PAVL3	P04863 canine parv
19	193	5.9	722	1 COAT_PAVL3	P24029 aleutian mi
20	186.5	5.7	647	1 COAT_PAVL3	P061575 mus musculu
21	143.5	4.4	648	1 COAT_PAVL3	P061575 mus musculu
22	123	3.8	648	1 COAT_PAVL3	P061575 mus musculu
23	123	3.8	648	1 COAT_PAVL3	P061575 mus musculu
24	121	3.7	880	1 COAT_PAVL3	P061575 mus musculu
25	121	3.7	880	1 COAT_PAVL3	P061575 mus musculu
26	119	3.6	667	1 COAT_PAVL3	P061575 mus musculu
27	116.5	3.6	642	1 COAT_PAVL3	P061575 mus musculu
28	110.5	3.4	655	1 COAT_PAVL3	P061575 mus musculu
29	110.5	3.4	1742	1 COAT_PAVL3	P061575 mus musculu
30	109.5	3.4	531	1 COAT_PAVL3	P061575 mus musculu
31	109.5	3.4	670	1 COAT_PAVL3	P061575 mus musculu
32	109	3.4	1186	1 COAT_PAVL3	P061575 mus musculu
33	108.5	3.3	860	1 COAT_PAVL3	P061575 mus musculu

34	108.5	3.3	1379	1 YFP9_SCHPO	O14066 schizosacch
35	108.5	3.3	1849	1 YFP9_SCHPO	P45386 haemophilus
36	107	3.3	2493	1 YFP9_SCHPO	P49606 usstilaigo ma
37	106.5	3.3	1070	1 Y355_HUMAN	O15063 homo sapien
38	106	3.3	1271	1 Y338_MYCGE	P47580 mycoplasma
39	105.5	3.2	559	1 HNF8_PIG	Q03365 sus scrofa
40	105	3.2	1396	1 ITA2_DROME	P12080 drosophila
41	105	3.2	1656	1 OMPB_RICJA	O06653 r outer mem
42	104.5	3.2	717	1 CLAI_ARATH	Q38854 arabidopsis
43	104.5	3.2	1122	1 ADPI_MYCGA	Q49379 mycoplasma
44	104	3.2	395	1 MAP1_YEAST	P41910 saccharomyc
45	103.5	3.2	1196	1 AMYB_PAEPO	P21543 paenibacill

ALIGNMENTS

RESULT 1	ID	COA3_AAV2	STANDARD	PRT	504 AA
AC	P03135				
DT	21-JUL-1986	(Rel. 01, Created)			
DT	21-JUL-1986	(Rel. 01, Last sequence update)			
DT	01-NOV-1991	(Rel. 20, Last annotation update)			
DE	Probable coat protein 3.				
OS	Adeno-associated virus 2 (AAV2).				
OC	Viruses; BBDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.				
OX	NCBI_TaxID=10804;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=83164299; PubMed=6300419;				
RA	Srivastava A., Lusby E.W., Berns K.I.;				
RT	"Nucleotide sequence and organization of the adeno-associated virus 2				
RT	genome.";				
RL	J. Virol. 45:555-564(1983).				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@ib-sib.ch">license@ib-sib.ch</a> ).				
CC	-----				
DR	EMBL; J01901; AAA42376.1; --				
DR	PIR; A03698; VCPV3A.				
DR	InterPro; IPR001403; Parvo_coat.				
DR	Pfam; PF00740; Parvo_coat; 1.				
KW	Coat protein.				
SQ	SEQUENCE 504 AA; 56366 MW; 75899B017052B6B2 CRC64;				
QY	Query Match	67.0%;	Score 2177.5;	DB 1;	Length 504;
QY	Best Local Similarity	82.5%;	Pred. No. 3.4e-137;		
QY	Matches 392;	Conservative 33;	Mismatches 47;	Indels 3;	Gaps 2;
QY	66 MASGGAPMADNNEGADGAGVNASGNWCHDSTWLGDRVITTTSTRTWALPTYNNHLYKQISS				125
QY	1 MATGSGAPMADNNEGADGAGVNASGNWCHDSTWLGDRVITTTSTRTWALPTYNNHLYKQISS				60
QY	126 ASTGASNDNHFGYSTPWPGFDFNRFCHFSPRDWQRLNNWGFPRPKRLNFKLFNIQVK				185
QY	61 QS-GASNDNHFGYSTPWPGFDFNRFCHFSPRDWQRLNNWGFPRPKRLNFKLFNIQVK				119
QY	186 EYTTNDGVTITIANNLSTVQVDSSEYQLPYVLGSAHOGCLPFPADVFMIPOYGYLTIN				245
QY	120 EYTTNDGVTITIANNLSTVQVDSSEYQLPYVLGSAHOGCLPFPADVFMIPOYGYLTIN				179
QY	246 NGSQAVGRSSFFCYLFYPSQMLRTGNNFTFSYTFEEVFPFSSYAHQSQSLDRMLNPLIDQY				305
QY	180 NGSQAVGRSSFFCYLFYPSQMLRTGNNFTFSYTFEEVFPFSSYAHQSQSLDRMLNPLIDQY				239
QY	306 LYYLNTQNSGSAQNKDLFPGRGSPAGMSVQPKNTLPPPCYRQQRVSKTKTDNNSNFT				365



Db 240 LYYLSRTNTPSGTTTQSRLQFSQAGASDIRDSRNLPGPCYRQQRVSKTSADNNNSEYS 299  
QY 366 WTGASKYNLNGRESIINPGTAMASHKDEDEKFEPMSCVMI FGKESAGASNTALDNVMTD 425  
Db 300 WTGATKYHLNGRDLVNP--AMASHKDEDEKFEPMSCVLI FGKQSEKTNVNI EKMITD 357  
QY 426 EEEIKATNPVATERFGTVAVNFQSSSTDPATGVDHAMGALPGMVMQDRDYYLQGPIMAKI 485  
Db 358 EEEIGTNPVATEYGYGSVSTNLQRGNRQAATADVNTQGVLPGMVMQDRDYYLQGPIMAKI 417  
QY 486 PHTDGHFHPSPMLMGFGFLKPNPPQILIKNTVPANPAPFASATKFASTTQYSTG 540  
Db 418 PHTDGHFHPSPMLMGFGFLKHPPOILLIKNTVPANPSTTFSAAKFASTTQYSTG 472

RESULT 2

COAT\_PAVBO STANDARD; PRT; 673 AA.  
ID COAT\_PAVBO . STANDARD; PRT; 673 AA.  
AC P07297; Q84374;  
DT 01-APR-1988 (Rel. 07, Created)  
DT 01-APR-1988 (Rel. 07, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Coat protein VP2 [Contains: Coat protein VP3].  
OS Bovine parvovirus (BPV).  
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.  
OX NCBI\_TaxID=10784;

RP SEQUENCE FROM N.A.  
RX MEDLINE=87061184; PubMed=3783814;  
RA Chen K.C., Shull B.C., Moses E.A., Lederman M., Stout E.R.,  
RA Bates R.C.;  
RT "Complete nucleotide sequence and genome organization of bovine parvovirus."  
RL J. Virol. 60:1085-1097(1986).  
CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
DR EMBL; M14363; AAB59848.1; -.  
DR EMBL; M14363; AAB59849.1; -.  
DR PIR; A26104; VCPVB5.  
DR HSSP; P30129; 4DPV.  
DR InterPro; IPR001403; Parvo\_coat.  
DR Pfam; PF00740; Parvo\_coat; 1.  
KW Coat protein; Glycoprotein.  
FT CHAIN 1 673 COAT PROTEIN VP2.  
FT CARBOHYD 138 673 COAT PROTEIN VP3.  
FT CARBOHYD 343 343 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 344 344 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 446 446 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 639 639 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT DOMAIN 163 187 GLY-RICH.  
SQ SEQUENCE 673 AA; 75103 MW; 5F244642B2214831 CRC64;

Query Match 17.1%; Score 556; DB 1; Length 673;  
Best Local Similarity 25.0%; Pred. No. 1.5e-29;  
Matches 160; Conservative 98; Mismatches 239; Indels 144; Gaps 19;

QY 1 TAPGKKRPVEGSPQEPDSSSGIGKTGQAPAKKRLNFGQTGDSSEVPDPLGEP--PATP 58  
Db 100 TSKGGRALKRKLYFARSNKGAKKANREPAPSTSNQONMEVSNIDIPNDEAGNQPIELATR 159  
QY 59 AAVGPTTMASGGGAPMADNNEGADVGNASGNMHCDSWLGDVITTSRTWALPTYNNH 118  
Db 160 SVVSGSVGGG-----RGSGGVGYSTGTGTTIFSENIVTKTROFICIDIKNGH 211  
QY 119 LYKOISSASTGASNDNHYFGYSTPWGYFDENRPHCHFSPRDWQRLINNNGFRPKRLNFK 178

Db 212 LYKS-EVINTGDTAHRQY-AITPWSYENFNQYSSHFSFNDWQHLVNDYERFRPKAMIVR 269  
QY 179 LENIQYKEVTINDGVTTIAN-LTSTQVPSDSEYQLPYLGSAHQCLPPFPADVEMTP 237  
Db 270 VYNLQIKQIMTDGAMGTVYNNDLTAGMHI FCDGDHRYPYVQHPWDDQCMPELPNSIWELP 329  
QY 238 QYGYLT-----NGSQAVGRS-----SFYCLEYFSPQMLRTGNNFTSYTFEEVPPHS 286  
Db 330 QYAYIPAPISVDNNTNTTVEEHLKGVPLMYLENSDHEVLNRG----- 373  
QY 287 SYAHSQSLDRLMNPLIDQYLYLNRTQN-----QSGSAQNKDLLFSRGSF----- 331  
Db 374 -----RIVRYIQLWRLMDRKQHHIQHASDVGSTGQKQKLLIQRKQPNKQRF 424  
QY 332 AGMSVQPKWLPGPCYRQQRVSKTKTDNNNSFTWTGASKYNLNGRESIINPGTAMASHK 391  
Db 425 QNAALRTSNMMSGP-----GIARGTHNATLQTSAGALVTWVT 462  
QY 392 DDEDKFFPMSCVM-----IFGKE-----SAGASNTALDNVMTDE 426  
Db 463 NGAD-----VSGVRAVRVGYSTDPYGGQQPESDLRLRYSASABEQQNPILEN----- 512  
QY 427 EEEKATNPVATERFGTVAVNFQSSSTDPATGVDHAMGALPGMVMQDRDYYLQGPIMAKIP 486  
Db 513 -----AARHTFTREARTKLITGNSGADGDYKEMWMLPNQWDSAPISRYNPIWVKVP 564  
QY 487 HTDGHFHPSPMLMGFGFLKPNPPQILIK--NTPVPANPAPFASATKFASTTQYSTGQSV 544  
Db 565 RVNRKTLTLDTDGSI PMSHPPTIFIKLARIPVPGNGD-----SFLNITYVTGQVSC 615  
QY 545 EIEWELOKENSKRWNPEVQYTSNYAKSANVD-FTVDNNGLY 584  
Db 616 EVVWEVEKGTKNWRPEYMH5---ATNMSVDAYTINNAGVY 653

RESULT 3

COAT\_PAVHB STANDARD; PRT; 781 AA.  
ID COAT\_PAVHB . STANDARD; PRT; 781 AA.  
AC P07299;  
DT 01-APR-1988 (Rel. 07, Created)  
DT 01-APR-1988 (Rel. 07, Last sequence update)  
DT 01-APR-1990 (Rel. 14, Last annotation update)  
DE Probable coat protein VP1.  
OS Human parvovirus B19.  
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Erythrovirus.  
OX NCBI\_TaxID=10798;  
RP SEQUENCE FROM N.A.  
RC STRAIN=Isolate AU;  
RX MEDLINE=86200451; PubMed=3701931;  
RA Shade R.O., Blundell M.C., Cotmore S.F., Tattersall P., Astell C.R.;  
RT "Nucleotide sequence and genome organization of human parvovirus B19 isolated from the serum of a child during aplastic crisis."  
RL J. Virol. 58:921-936(1986).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
DR EMBL; M13178; AAA66867.1; -.  
DR PIR; A24299; VCPV19.  
DR InterPro; IPR001403; Parvo\_coat.  
DR Pfam; PF00740; Parvo\_coat; 1.  
KW Coat protein; Glycoprotein.  
FT CARBOHYD 46 46 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 184 184 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 220 220 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 293 293 N-LINKED (GLCNAC. . .) (POTENTIAL).



QY 582 GLYTEPRPIG 591  
DB 724 GKVI-PTNIG 732

## RESULT 5

COAL\_PAVP9 STANDARD; PRT; 729 AA.

AC P33484;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Coat protein VP1 [Contains: Coat protein VP2].  
OS Porcine parvovirus (strain 90HS) (PPV).  
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.  
OX NCBI\_Taxid=33725;

RN [1]  
RP SEQUENCE FROM N.A.

RX MEDLINE=89319168; PubMed=2750278;

RA Sakurai M., Nishimori T., Ushimi C., Nakajima H.;

RT "Nucleotide sequence of capsid protein gene of porcine parvovirus.";  
RL Virus Res. 13:79-86(1989).

CC -1- SUBUNIT: MATURE VIRION CONTAINS THREE CAPSID PROTEINS DESIGNATED  
VP1, VP2, AND VP3 AND A NONCAPSID PROTEIN NS-1.

CC -1- MISCELLANEOUS: VP3 MIGHT BE A POSTTRANSLATIONAL CLEAVAGE PRODUCT  
OF VP2 IN SEVERAL AUTONOMOUS PARVOVIRUSES.

CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.

DR PIR; A60006; A60006.

DR HSSP; P30129; 4DPV.

DR InterPro; IPR001403; Parvo\_coat.

DR Pfam; PF00740; Parvo\_coat; 1.

KW Coat protein; Glycoprotein.

FT CHAIN 1 729 COAT PROTEIN VP1.

FT CHAIN 1 729 COAT PROTEIN VP2.

FT CARBOHYD 172 172 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 198 198 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 282 282 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 433 433 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 471 471 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 573 573 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 604 604 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 651 651 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 651 651 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 651 651 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 651 651 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 651 651 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 651 651 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 651 651 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 651 651 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 651 651 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 651 651 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 651 651 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 651 651 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 651 651 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 651 651 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 651 651 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 651 651 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 651 651 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 651 651 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 651 651 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 651 651 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 651 651 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 651 651 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 651 651 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 651 651 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 651 651 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 651 651 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 651 651 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 651 651 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 651 651 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 651 651 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 651 651 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 651 651 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 651 651 N-LINKED (GLCNAC. . .) (POTENTIAL).

QY 314 NQSGAONKDLLFSRGSPPAGMSVQPKNWLPGPCYRQQRVSKTKTDNNNSFTWTGASKYN 373  
DB 462 TRKGYHQTMNNSYTEAT-AIRP-----AQVGYNTPYMFEYSNGGPF- 502

QY 374 LINGRESINP--GTAMASHKDEDEKFFPMGSMVIFGKESAGASNTALDNVMTDEEIIKA 431  
DB 503 -----LTPIVPTADTQYNDDE-----PNGAIRFTMGYQHGLTTS-----SQELERYT 545

QY 432 TNP-----VATERFGTVA-VNEQSSS-----TDPATG--DVHAMGALP-----GMYWQ 471  
DB 546 FNPQSKGRAPKQGFNQAPLNLNTNNGTLLPSDPFGKPMVHFMNTLNTYGPLTALNN 605

QY 472 DRDVLQGPPIWAKIPHTD--GHFHSPLMGFGGLK-NPPQILIKNTVPVAPNPAEFSA- 527  
DB 606 TAPVFPNGQIWDKELDTLKPRLH--VTAPFVCKNPNPGQLFVKIAP--NLTDDEFNAD 659

QY 528 TKRASFITQYSTGQVSVEIEMELQKNSKRNPEVOYTSNYSKASANDFTVDNGLYTEP 587  
DB 660 SPQOPRIITYSNFWWKGTLTFTAKMRSSNMNPIQOHTT-----TAENIGNYI-P 708

QY 588 RPIG 591  
DB 709 TNIG 712

QY 588 RPIG 591  
DB 709 TNIG 712

QY 588 RPIG 591  
DB 709 TNIG 712

QY 588 RPIG 591  
DB 709 TNIG 712

QY 588 RPIG 591  
DB 709 TNIG 712

QY 588 RPIG 591  
DB 709 TNIG 712

QY 588 RPIG 591  
DB 709 TNIG 712

QY 588 RPIG 591  
DB 709 TNIG 712

QY 588 RPIG 591  
DB 709 TNIG 712

QY 588 RPIG 591  
DB 709 TNIG 712

QY 588 RPIG 591  
DB 709 TNIG 712

QY 588 RPIG 591  
DB 709 TNIG 712

QY 588 RPIG 591  
DB 709 TNIG 712

QY 588 RPIG 591  
DB 709 TNIG 712

QY 588 RPIG 591  
DB 709 TNIG 712

QY 588 RPIG 591  
DB 709 TNIG 712

QY 588 RPIG 591  
DB 709 TNIG 712

QY 588 RPIG 591  
DB 709 TNIG 712

QY 588 RPIG 591  
DB 709 TNIG 712

QY 588 RPIG 591  
DB 709 TNIG 712

QY 588 RPIG 591  
DB 709 TNIG 712

QY 588 RPIG 591  
DB 709 TNIG 712

QY 588 RPIG 591  
DB 709 TNIG 712

QY 588 RPIG 591  
DB 709 TNIG 712

QY 588 RPIG 591  
DB 709 TNIG 712



FT	CHAIN	1	729	COAT PROTEIN VP1.
FT	CHAIN	151	729	COAT PROTEIN VP2.
FT	CARBOHYD	172	172	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	198	198	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	282	282	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	330	330	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	433	433	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	471	471	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	573	573	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	604	604	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	651	651	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	DOMAIN	175	190	GLY-RICH.
SO	SEQUENCE	729 AA;	80946 MW;	EF816E246C80DB42 CRC64;

Query Match 7.8%; Score 254.5; DB 1; Length 729;  
Best Local Similarity 22.7%; Pred. No. 1.6e-09;  
Matches 152; Conservative 94; Mismatches 272; Indels 151; Gaps 34;

Qy	9	VEQSPQEPDSSSGIGK-----TGQQPAKKRLNFGQTGDSSEVPDPQPLGEPATP	58
Db	109	VRSPRKHGSKPPGKRPA RHIFINLAKKAKGTSNTNSMSSENVQHNPI NAGTEL-	167
Qy	59	AAVGPTTMA SGGAPMADNNEGADGVNASG--NMHCDSTWLG D--RVITSTRTWALP	113
Db	168	SATGNESGGGGGGG---GGRGAGGVSTGTENNQTEFOYLGEGLVRITAHASRLIH L N	223
Qy	114	TYNNHLKQIS--SASTGAS---NDNHYFGYSTPWGYFDENRFHCHFS PRDWQRLINN	167
Db	224	MPEHETRYKRIHVLNSES SVAGQWQDDAHTQMTWPMSLIDANAMGWENPADQOLISNM	283
Qy	168	WGFRPKRLNFKLFNIQVKEV T--TNDGVTTIANNLSTVOVFSSEYQLPYVLGSAHOG	224
Db	284	TEINLVSFEOEIFNVVLKTTESATSPPTKIYNNDLTASLMVALDTNNTLPYTPAAPRSE	343
Qy	225	CLPPPADVFMIPQGYL-----TLNNGSQAVGRS-----SFYCLE-YFPS	264
Db	344	TLGFYFWLPFKPTQYRYLLSCIRNLNPPYTGQSQQITDSIQTGLHSDIMFYTIENAVPI	403
Qy	265	QMLRTGNNE-TFSYTFEEV PFHSSYAHQSQSLDR-----LMNP LI--DOYLYYLNRTQ	313
Db	404	HLRTGDEFSTGIYHFDTKPL--KLTHSWQTNRS LGLPKLLTEPTTEGDQHPTLPAAN	461
Qy	314	NQSGAQNKDLLFSRSGPAGMSVQPKNMLBGP CYRQRVSKTKTDNNNSNFTWTGASKVN	373
Db	462	TRKGYHQTINNSYTEAT---AIRP-----AQVGYNTPYMNFEYSNGGPF-	502
Qy	374	LNGRESIINP--GTAMASHKDDDEKFFPM SGVMI FGKESAGASNTALD--NVMITDEE	428
Db	503	-----LTPIVPTADTQYNDDEPN-----GAIRFTMDYQHGHLTSSQE	540
Qy	429	IK--ATNP-----VATERFGTVA-VNQSSS-----TDPATG--DVHAMGALP----	466
Db	541	LERYTFNPQSKCGRAPKQOFNQAPLNLBENTNNGTL LPSDPIGKSNMHFMNTLNTYGPL	600
Qy	467	GMVWQDRDYVLQGP I WAKI PHTD--GHFHPSP LMGFG LK--NPPQILIKNTPV PANPPA	523
Db	601	TALNNTAPYFENQIWDKELDTLKPRLH--VTAPFVCKNNPPGQLFVKIAP--NLTD	654
Qy	524	EFSA--TKFASFITQYSTGVSEI EWELOKENS KRMNPEVOYTSNVAKSANVDFTVDNNG	582
Db	655	DFNADSPQQRILITYSNFWWKGLTFTAKGRSSNMNMNPIQOHTT-----TAENIG	704
Qy	583	LYTEPRPIG 591	
Db	705	NYI-PTNIG 712	

RESULT 7  
COAL\_PAVPK STANDARD; PRT; 729 AA.  
AC P52501;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT	16-OCT-2001	(Rel. 40, Last annotation update)
DE	Coat protein VP1	{contains: Coat protein VP2}.
OS	Porcine parvovirus (strain Kresse) (PPV).	
OC	viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.	
OX	NCBI_Taxid=73487;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=96183900; PubMed=8642680;	
RA	Bergerson J., Hebert B., Tijssen P.;	
RT	"Genome organization of the Kresse strain of porcine parvovirus:	
RT	identification of the allotropic determinant and comparison with	
RT	those of MADL-2 and field isolates.";	
RL	J. Virol. 70:2508-2515(1996).	
CC	-I- SUBUNIT: MATURE VIRION CONTAINS THREE CAPSID PROTEINS DESIGNATED	
CC	VP1, VP2, AND VP3 AND A NONCAPSID PROTEIN NS-1.	
CC	-I- MISCELLANEOUS: VP3 MIGHT BE A POSTTRANSLATIONAL CLEAVAGE PRODUCT	
CC	OF VP2 IN SEVERAL AUTONOMOUS PARVOVIRUSES.	
CC	-I- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>	
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).	
CC	-----	
DR	EMBL; U44978; AAC40230.1; -.	
DR	EMBL; U44978; AAC40231.1; -.	
DR	HSSP; P30129; 4DPV.	
DR	InterPro; IPR001403; Parvo_coat.	
DR	Pfam; PF00740; Parvo_coat; 1.	
KW	Coat protein; Glycoprotein.	
FT	CHAIN	1 729
FT	CHAIN	151 729
FT	CARBOHYD	172 172
FT	CARBOHYD	198 198
FT	CARBOHYD	282 282
FT	CARBOHYD	330 330
FT	CARBOHYD	433 433
FT	CARBOHYD	471 471
FT	CARBOHYD	573 573
FT	CARBOHYD	604 604
FT	CARBOHYD	651 651
SO	SEQUENCE	729 AA; 80835 MW; B6345BFA0568A1F6 CRC64;

Query Match 7.8%; Score 252.5; DB 1; Length 729;  
Best Local Similarity 22.7%; Pred. No. 2.2e-09;  
Matches 151; Conservative 95; Mismatches 277; Indels 141; Gaps 33;

Qy	9	VEQSPQEPDSSSGIGK-----TGQQPAKKRLNFGQTGDSSEVPDPQPLGEPATP	58
Db	109	VRSPRKHGSKPPGKRPA RHIFINLAKKAKGTSNTNSMSSENVQHNPI NAGTEL-	167
Qy	59	AAVGPTTMA SGGAPMADNNEGADGVNASGNM--CDSTWLG D--RVITSTRTWALP	113
Db	168	SATGNESGGGGGGG---GGRGAGGVSTGTFNNQTEFOYLGEGLVRITAHASRLIH L N	223
Qy	114	TYNNHLKQIS--SASTGAS---NDNHYFGYSTPWGYFDENRFHCHFS PRDWQRLINN	167
Db	224	MPEHETRYKRIHVLNSES SVAGQWQDDAHTQMTWPMSLIDANAMGWENPADQOLISNM	283
Qy	168	WGFRPKRLNFKLFNIQVKEV T--TNDGVTTIANNLSTVOVFSSEYQLPYVLGSAHOG	224
Db	284	TEINLVSFEOEIFNVVLKTTESATSPPTKIYNNDLTASLMVALDTNNTLPYTPAAPRSE	343
Qy	225	CLPPPADVFMIPQGYL-----TLNNGSQAVGRS-----SFYCLE-YFPS	264
Db	344	TLGFYFWLPFKPTQYRYLLSCIRNLNPPYTGQSQQITDSIQTGLHSDIMFYTIENAVPI	403
Qy	265	QMLRTGNNE-TFSYTFEEV PFHSSYAHQSQSLDR-----LMNP LI--DOYLYYLNRTQ	313
Db	404	HLRTGDEFSTGIYHFDTKPL--KLTHSWQTNRS LGLPKLLTEPTTEGDQHPTLPAAN	461

QY 314 NOSGSAQNKDLFSGSPAGMSVOPKMWLPQPCYRQQRVSKTKTDNNNSNFTWTGASKYN 373  
DB 462 TRKGHYQTINNSYTEAT---AIRP-----AQVGYNTPYMFYSNGGPF- 502  
QY 374 LNGRESIINP--GTAMASHKDEDEKFFPMGVMIFGKESAGASNTALDNVMTDEEIIKA 431  
DB 503 -----LTPIVPDTADTQYNDDE---PNCAIRFTMGYQHGLTTS-----SQELERYT 545  
QY 432 TNP-----VATERFGTVA-VNFQSSS-----TDPATG--DVHAMGALP-----GMVWQ 471  
DB 546 FNPQSKCGRAPKQGFNQAPLNLENTNNGTLLPSDPIGKPMHFMNTLNTYGPILTALNN 605  
QY 472 DRDVTYLGPIWAKIPHTD--GHFHPSPLMGFGFLK-NPPQQLIKNTPYPANPPAEFSA- 527  
DB 606 TAPVFPNGQIWDEKLDLTKPRLH---VTAPFVCKNNPPGQLVYKIAP---NLTDDEFNAD 659  
QY 528 TKFASFTQYSTGVSVSEIEMELQKENSKRMPDEVQYTSNYAKSANVDFTVDNNGLYTEP 587  
DB 660 SPQQRITITSNFWWKGTLLTFTAKMRSSNMNPIQOHTT-----TAENIGNYI-P 708  
QY 588 RPIG 591  
DB 709 TNIG 712

RESULT 8  
COAT\_PAVHH STANDARD; PRT; 722 AA.  
ID COAT\_PAVHH STANDARD; PRT; 722 AA.  
AC P03136;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Coat protein VP1 [Contains: Coat protein VP2].  
OS Hamster parvovirus H1.  
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.  
OX NCBI\_TaxID=10799;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=83112183; PubMed=6823009;  
RA Rhode S.L. III, Paradiso P.R.;  
RT "Parvovirus genome: nucleotide sequence of H-1 and mapping of its  
RT genes by hybrid-arrested translation."  
RL J. Virol. 45:173-184(1983).  
CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL; X01457; CAB57285.1; ALT\_SEQ.  
DR PIR; A03699; VCPVW2.  
DR HSSP; P07302; IMVM.  
DR InterPro; IPR001403; Parvo\_coat.  
DR Pfam; PF00740; Parvo\_coat; 1.  
KW Coat protein; Glycoprotein.  
FT CHAIN 1 722 COAT PROTEIN VP1.  
FT CHAIN 131 722 COAT PROTEIN VP2.  
FT CARBOHYD 178 178 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 218 218 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 639 639 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 647 647 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 675 675 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT DOMAIN 155 170 GLY-RICH.  
SQ SEQUENCE 722 AA; 79737 MW; 6BB678391A5DC31 CRC64;  
Query Match 7.5%; Score 244; DB 1; Length 722;  
Best Local Similarity 21.4%; Pred. No. 8e-09;  
Matches 142; Conservative 94; Mismatches 267; Indels 160; Gaps 29;

QY 8 PVEQSPQEPDSSSGIGKTQQ---PA-----KKRLNFGQTGDSSEVPDPQPLGE 54  
DB 82 PKLSTUSEP-GTSGVSRPGKRTKPAHIFVNQARAKKRLASLAAQOQRTLTMSDGETNQ 140  
QY 55 PATPAVGFPTMASGGAPMADNEGADGVNAGSNWCHDSTW--LGDRVITSTRITWAL 112  
DB 141 DTGIANARVERADGGG-SGGGGSGGGGIGVSTGYDNOQTYKELGDGWEITAHASRL 199  
QY 113 -----PTYNHLYKQISSASTGASNDNHFGYSTPWGYFDNFRFCHFSR 158  
DB 200 LHLGMPSENYCRVTVHNQTTGHTKVKGNMAYDTHQIW--TPWSLVDANAMGWVFPQS 258  
QY 159 DWQRLNNNMWGERPKRLNFKLFNIQVEVTT-----NDGVTTIANNLSTVQVFSDEYQ 213  
DB 259 DWQFIQNSMESLNLDSLQELFNVYKVTVEQAGQDAIKYNNNDLTACMMVALDSNNI 318  
QY 214 LPYVLGSAHQGL-----PPFPAD---VEMIPQYGLTLNNGSQ-----AVG-----R 253  
DB 319 LPYTPAAQTSETLGFPWKPTAPAPRYFFFMPRQLSVTSSNSAEGTQITDTIGEPOALN 378  
QY 254 SSFYCLE-YFPSQMLRTGNFTF-SYTFEEVPHSSYA-----HSQSLDRLMN-PLIDQY 305  
DB 379 SQFFTIENTLLPITLRLTGDEFTTGTYIENTDPLKLTHTWQNRHLACLQGITDLPSTDA 438  
QY 306 LYYLNRTONQSGSAQNKDLF-----SRGSPAGMSVOPKMWL 342  
DB 439 TASTLTANGDRFGSTQTONVNYTEALRTPAQIGFMQPHDNEANRGPFKVPVVP---- 494  
QY 343 PGPCYRQQRVSKTKTDNNNSNFTWTGASKYNLNGR--ESLNPQTA-----MASH 390  
DB 495 -----LDITAGEDHDAN---GAIRFNYGKHGEDWAKQGAAPERWTDAIDSAAG 541  
QY 391 KDEDKFFPMGVMIFGKESAGASNTALDNVMTDEEIIKA-TNPVATERFGTVAVNPOS 449  
DB 542 RDTARCFV-----QSAPISIPPNQNLQREDALAGRNMHYTNVFNSTYGPLSAF 591  
QY 450 SSTDPATGDVHAMGALPGMWQDRDVTYLGPIWAK---IPHTDGHFHPSPLMGFGFLKNP 506  
DB 592 PHPDP-----IYNGQIWDEKLDLEHKRLHVTAPV---CKNNP 628  
QY 507 PPQILLIKNTPYPANPPAEF-SATKFASTITQYSTGVSVSEIEMELQKENSKRMPDEVQY 564  
DB 629 PGQLFVHLGP--NLTDQFDPNSTTVSRIVT-YSTFYWKGLIKFAKAKLRPLTNPNVYQA 684  
QY 565 TSN 567  
DB 685 TTD 687

RESULT 9  
COAT\_PAVL3 STANDARD; PRT; 587 AA.  
ID COAT\_PAVL3 STANDARD; PRT; 587 AA.  
AC P36310;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Coat protein VP1 [Contains: Coat protein VP2].  
OS Parvovirus LuIII.  
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.  
OX NCBI\_TaxID=35339;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93297126; PubMed=8517025;  
RA Diffoot N., Chen K.C., Bates R.C., Lederma M.;  
RT "The complete nucleotide sequence of parvovirus LuIII and  
RT localization of a unique sequence possibly responsible for its  
RT encapsidation pattern."  
RL Virology 192:339-345(1993).  
CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
DR EMBL; M81888; -; NOT\_ANNOTATED\_CDS.  
DR PIR; B44276; B44276.  
DR HSSP; P07302; 1MWV.  
DR InterPro; IPR001403; Parvo\_coat.  
DR Pfam; PF00740; Parvo\_coat; 1.  
DR Coat protein; Glycoprotein.  
KW CARBOHYD 49 49  
FT CARBOHYD 90 90 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 220 220 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 304 304 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 371 371 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 503 503 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 511 511 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 514 514 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 539 539 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 571 571 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 587 AA; 65429 MW; 523E6B9CBF2EBB74 CRC64;

Query Match 7.2%; Score 234; DB 1; Length 587;  
Best Local Similarity 23.1%; Pred. No. 2.8e-08;  
Matches 137; Conservative 74; Mismatches 231; Indels 152; Gaps 29;  
OY 59 AAVGPTTASGGGAPMADNNEGADVGNASGNWHCHDS--TWLGD--RVITSTRWAL- 112  
DB 23 AADGPG--GSGGG-----GSGGGGVSSTGSDYNQTHYKFLGDGWEITAYSTRMHLN 74  
OY 113 -PTYN-----HLKYISSASTGASNDNHYFGYSTPWGYFDNFRHGHFSPRDMORLIN 165  
DB 75 MPKSENYCRVRVNTNDGTASHMAMDHEQIW-TPSLVDANAMGWVFPSPDWQYISN 133  
OY 166 NNWGFPRKLNFKLFNIQVKEVT-TNDGVTTIA--NMLSTVOVESDSEYQLPYVLGSA 221  
DB 134 NMHINHLSDQELFNVIKTVEQNTGAERIKVYNNDLTAAMVVALDSNNILPYTPAID 193  
OY 222 HQGCL-----PPFPA-----DYFMIPOYGYLTLNNGSQAVGRSSFYCLEYFP 263  
DB 194 NQETLGFYPMKPTIPSPRYFYFSCDRNLSTYKDEAGTITDTMGLASGLNSQFTIENTQ 253  
OY 264 S-QMLRTGNF-TFSYTEEVPHSSYAHQSLSRLMPLIDQ-----YL 306  
DB 254 RINLRTGDEYATGTYTFDTEPIRLTHWTQNRHLGQRPQITELPSSDTANATLTARGYR 313  
OY 307 YLLNRTQN-----QSGSAQNKD-LIFSRGSPAGMSVQPKWLPG----- 344  
DB 314 SGLTQIGRNDVTEATRVRAQVGFQCPHNFETSRAGPFKVPVVPADITQGLDHANGS 373  
OY 345 --PCYRQQRVSKTKTDNNNSNFTWTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMG 402  
DB 374 LRYTYDKQHGQSWASQNNKDRYTW-DAVNYD-SGR----- 406  
OY 403 VMIFGKESAGASNTALDNVMTDEEEIKATNPVATER--FGTAVNFGS--SSTDPATG 457  
DB 407 -----WTNNCFIQSVPTSEPN--ANQILTNRDNLAKGTDHFTNAFNSYGPLTA 454  
OY 458 DVHAMGALPGWVQDRDYLQSPIWAK---IPHTDGHFHPSPLMGFGKKNPPQILIKN 514  
DB 455 FPH-----PAPYPOGQIWDKELDLHKKPRLTQAPFV---CKNNAPGQLLVRL 500  
OY 515 TP--VPANPAEFSATKFAFITQYSTQVSEIEMELQKENSKRNNDEVQYTS 566  
DB 501 APNLTQDYDNNSSNLSRIVTYGTFFWKGLTLKAK--MRPNA-TWNPVFQISA 550

RESULT 10  
COAT\_MUMIV  
ID COAT\_MUMIV STANDARD; PRT; 716 AA.  
AC P03137;

DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, last sequence update)  
DT 16-OCT-2001 (Rel. 40, last annotation update)  
DE Coat protein VP1 [Contains: Coat protein VP2].  
OS Murine minute virus (Murine parvovirus).  
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.  
OX NCBI\_TaxID=10794;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=83143341; PubMed=6298737;  
RA Astell C.R., Thomson M., Merchlinsky M., Ward D.C.;  
RT "The complete DNA sequence of minute virus of mice, an autonomous  
RT parvovirus.";  
RL Nucleic Acids Res. 11:999-1018(1983).  
CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
DR EMBL; V01115; CAA24310.1; ALT\_SEQ.  
DR PIR; A03700; VCPV2M.  
DR HSSP; P07302; 1MWV.  
DR InterPro; IPR001403; Parvo\_coat.  
DR Pfam; PF00740; Parvo\_coat; 1.  
DR Coat protein; Glycoprotein.  
KW CHAIN 1 716  
FT CHAIN 1 716  
FT CARBOHYD 131 716 COAT PROTEIN VP1.  
FT CARBOHYD 179 716 COAT PROTEIN VP2.  
FT CARBOHYD 218 716 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 218 218 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 500 500 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 633 633 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 669 669 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT DOMAIN 156 171 GLY-RICH.  
SQ SEQUENCE 716 AA; 78707 MW; 6A7229A91161F4C6 CRC64;

Query Match 6.6%; Score 216; DB 1; Length 716;  
Best Local Similarity 20.8%; Pred. No. 5.7e-07;  
Matches 148; Conservative 93; Mismatches 271; Indels 198; Gaps 34;

OY 8 PVEQSPQEPDSSSGIGKTG-----QQPAKKRL-----NFGQT-GDSESVPDP 48  
DB 82 PKLATDSEF-GTSGVSRAGKRTTPPAYIFINQARAKKLTSSAAQSSQTMDSGTSQPS 140  
OY 49 QPLGEPPA--TPAAVPTTASGGGAPMADNNEGADVGNASGNWHCHDS--TWLGD--R 101  
DB 141 GNAVHASARVERAADPG--GSGGG-----GSGGGGVSSTGSDYNQTHYRFLGDGWE 192  
OY 102 VITGSTRWALPTYNHLYKQI-----SSASTGASNDNHYFGYSTPWGYFDNFRHCH 154  
DB 193 ITALATRLVHLNMPKSENYCRIRVHNTTDSYKGNMAKDAHEQIWTPSLVDANAMGVW 252  
OY 155 FSPRDMQRLINNNGWFRPKRLNFKLFNIQVKEVTND---GVTIANNLSTVOVFSDS 210  
DB 253 LQPSDWQYICNTMSQNLVSLDQELFNVLKVTBQDLGQAIKIYNNDLTACMMVAVDS 312  
OY 211 EYQLPYVLGSAHQGLPPFPADVFMIPQYGY-----LTNNGSQAV-----G 252  
DB 313 NNILPYTPAANSMETLGFYPMKPTIASPYRYFCVDRDLSTYENQEGTVEHNVMGTPKG 372  
OY 253 RSSGFCLYFPPS-QMLRTGNF-TFSYTEEVPHSSYAHQSLSRLMPLIDQYLYY 308  
DB 373 IPOFTIENTQITLRTGDEYATGTYTFDTSV--KLTHWTQNRQLGQPLSTF--- 427  
OY 309 LNRQNSGSAQNKDLLFSRGSBAGMSVQPKWLP-----PG 344  
DB 428 -PEADTDAGT-----LTAQGSRHGTTQMGVNVWSEAIRTPRAQVGFQCPHNDFEASRAG 480  
OY 345 PCYRQQRVSKTKTDNNNSNFTWTGASKYNLNGRESIINPG-----TAMASHKDE 394





[3]  
RN X-RAY CRYSTALLOGRAPHY (3.5 ANGSTROMS) OF 132-718.  
RA lllamas-Saiz A.L., Agbandje-Mckenna M., Wikoff W.R., Bratton J.,  
RA Tattersall P., Rossmann M.G.;  
RT "Structure determination of Minute Virus of mice.";  
RL Acta Crystallogr. D 53:93-100(1997).  
CC -!- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X02481; CAB46507.1; -.  
DR EMBL; X02481; CAB46508.1; -.  
DR EMBL; M12032; AAA69569.1; ALT\_INIT.  
DR PIR; B23008; VCPVIM.  
DR PDB; 1MVM; 25-FEB-98.  
DR InterPro; IPR001403; Parvo\_coat.  
DR Pfam; PF00740; Parvo\_coat; 1.  
KW Coat protein; Glycoprotein; 3D-structure.  
FT CHAIN 1 718 COAT PROTEIN VP1.  
FT CHAIN 132 718 COAT PROTEIN VP2.  
FT CARBOHYD 180 180 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 219 219 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 502 502 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 635 635 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 671 671 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT DOMAIN 157 172 GLY-RICH.  
FT CONFLICT 144 144 A -> G (IN REF. 2).  
SQ SEQUENCE 718 AA; 79017 MW; B43C1762ED5F74B9 CRC64;  
  
Query Match 6.6%; Score 214.5; DB 1; Length 718;  
Best Local Similarity 21.0%; Pred. No. 7.1e-07;  
Matches 142; Conservative 89; Mismatches 273; Indels 171; Gaps 30;  
  
QY 8 PVEGSGQEPDSSSGIGKTG-----QQPAKRLNFG--QTGDSSEVPDPPLG 52  
DB 83 PKLATDSEP-GTSGVSRAGKRTTPPAYIFINQARAKKLTSSAAQSSQTMDSGTSDPDG 141  
QY 53 -----EPPATPAVGPPTTMASGGAPMADNNEGADGVNAGSNWCHDS--TWLGD--R 101  
DB 142 GNAVHSAKVERADPG--GSGGG-----GSGGGGVGVSTGSYDNQTHYRFLGDGWVE 193  
QY 102 VITTSRRTWALPTYNHLYKQI-----SSASTGASNDNHFGYSTPWGXFDFNRHCH 154  
DB 194 ITALATRLVHLNMPKSENYCRIRVHNTDTSVKGNMAKDADHEQITWPSLV DANAMGVW 253  
QY 155 FSPRDWQRLINNNGFRPKRLNFKLFNIOVKEVTND---GVTTIANNLSTVQVFSDS 210  
DB 254 LQPSDWQYICNTMSQNLVSLDQEIFNVVLKVTBQDSGGQAIKTYNNDLTACMMVAVDS 313  
QY 211 EYQLPYVLGSAHQCLPPFPADVFEMIPQY-----LTLNN-----GSQAV 251  
DB 314 NNILPYTPAANSMETLGFYPWKPTIASPRYYFCVDRDLSTYENQEGTIEHNVMGTPKG 373  
QY 252 GRSSFYCLEYFPPS-QMLRTGNNF-TFSYTFEEVFPFHSSYAHSGSLRLMN--PLIDQYLY 307  
DB 374 MNSQFFTIENQITLLRTGDEFATGYFDTPNV--KLTHWQTNRQLGQPLSTF-- 429  
QY 308 YLNRTQNGSGAQNKDLFSRGSPPAGMSVQPKNWL-----P 343  
DB 430 --PEADTDAGT-----LTAQGSRHGATQMEVNMVWSEAIRTPAOGVFCQPHNFEASRA 481  
QY 344 GPCYRQGRVSKTKTDNNNSFTWTGASKYNLNGRESIINPGTAMASHKDEDEK----- 397  
DB 482 GP-FAAPKVPADVTOGVDRANGSVRSYSGKHGEMWAAHGPAPERITYWDETDFGSGRDT 540  
QY 398 ---FPMGVMIFGKESAGASNTALDNVWITDDEEIKATNPVATERFGTAVAVNFOS--SST 452

DB 541 RDGFIQSAPLVPPPLNGILTNA-----NPIGTKN---DIHFSNVENS 581  
QY 453 DPATGDVHAMGALPGMWQDRDYVLQGPIMAK---IPHTDGHFHPSPLMGGFGLKNPPQ 509  
DB 582 GPLTAFSH-----PSPYYPQGIWDKELDLHKRDLHITAPFV---CKNNAPGQ 627  
QY 510 ILIKNTP--VPANPAEFSAFKFASFITOYSTGQSVSEIEMELQENSKRWNPVEQYTSN 567  
DB 628 MLYRLGPNLTDQYDPNGATLSRIYVTGYTFWKGKLTMRALRA-----NTWNPVYQ---- 679  
QY 568 YAKSANVDFTVDNG 582  
DB 680 -----VSVEDNG 686  
  
RESULT 13  
COAT MEVA  
ID \_COAT\_MEVA STANDARD; PRT; 722 AA.  
AC P27437;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Coat protein VP1 [Contains: Coat protein VP2].  
OS Mink enteritis virus (strain Abashiri) (MEV).  
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.  
OX NCBI\_TaxID=10793;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91202123; PubMed=2016597;  
RA Kariatsumari T., Horiuchi M., Hama E., Yaguchi K., Ishiguro N.,  
RA Goto H., Shinagawa M.;  
RT "Construction and nucleotide sequence analysis of an infectious DNA  
RT clone of the autonomous parvovirus, mink enteritis virus.";  
RL J. Gen. Virol. 72:867-875(1991).  
CC -!- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; D00765; BAA00663.1; -.  
DR PIR; B38350; VCPVME.  
DR HSSP; P30129; 4DPV.  
DR InterPro; IPR001403; Parvo\_coat.  
DR Pfam; PF00740; Parvo\_coat; 1.  
KW Coat protein; Glycoprotein.  
FT CHAIN 1 722 COAT PROTEIN VP1.  
FT CHAIN 139 722 COAT PROTEIN VP2.  
FT DOMAIN 160 177 GLY-RICH.  
FT CARBOHYD 163 163 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 185 185 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 202 202 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 318 318 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 581 581 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 643 643 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 655 655 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 722 AA; 79823 MW; 9DADBACAB2EF9F622 CRC64;  
  
Query Match 6.6%; Score 214; DB 1; Length 722;  
Best Local Similarity 20.2%; Pred. No. 7.8e-07;  
Matches 139; Conservative 101; Mismatches 259; Indels 188; Gaps 34;  
  
QY 7 RPVEGSGQEPDSSSGIGKTGQPAKRLNFGQTGDSSEVP---DPQLGEPPA--TPAA 60  
DB 105 KPTKRSKPPPHIFINLAK-----KKAGAGQVQRDNLAPMSDGA VQPDGQPAVRNERA 158  
QY 61 VGPTTMASGGAPMADNNEGADGVNAGSNW-----CDSTWLGDRTVITSTRTWALP 113  
DB 159 TSGSGNGSGGG-----GGSGGVGISTGTFTNNQTEPKFLENGWV--EITANSRLVHLN 210

```

QY      114 TYNNHLYKQI-----SSASTGASNDNHFGYSTPWGYFDENRFCHFSPRDQRLINN 166
        : ||::: : : : : ||| | : | : | : | : | : | : |
Db      211 MPESENYKRVRVNNMDKTAVKGMMALDDTHQIVTFWSLVDANAMGWFWFNPGDWQLIVNT 270
        : ||::: : : : : ||| | : | : | : | : | : | : |
QY      167 NWGFRRPKRLNFKLFNIQWKEVT---TNDGVTIANNLSTVQVFSDEYQLPYVLGSAHQ 223
        : ||::: : : : : ||| | : | : | : | : | : | : |
Db      271 MSEHLVSFEQEIEFNVLKTVESASATQPRTKYANNDLTASLMVALDSNNTMPTFPAAIRS 330
        : ||::: : : : : ||| | : | : | : | : | : | : |
QY      224 GCL-----PPFPA-----DVFMIPQYGXLTLLNGSQAVGRS-----SFYCL 259
        | | | | | | | | | | | | | | | | | | | | | | |
Db      331 ETLGEYPWKPTIPTPMRYYFQWDRTLIPSH-----TGTSGETINIYHGTDPDQFYTI 384
        : ||::: : : : : ||| | : | : | : | : | : | : |
QY      260 E-YFPSOMLRGTGNF-TFSYTBEVPFHSSIAHSQSLDRLMN--PLIDLYLYLNRTQNO 315
        - : ||::: : : : : ||| | : | : | : | : | : | : |
Db      385 ENSGVPHLLRTGDDEFATGTFFEDCKP-CRLTHWTQNRALGLP-----FLNSLPQS 435
        : ||::: : : : : ||| | : | : | : | : | : | : |
QY      316 SGSQNKDLLFSRGSPAGMS-----VQPKN-WLPGPCYRQRVSK----- 354
        : | : | : | : | : | : | : | : | : | : | : |
Db      436 EGATNFGDIGVOQDKRGVTOMGNTDYITEAIRPAEVGSAPYSSF EASTOGPFETPI 495
        : ||::: : : : : ||| | : | : | : | : | : | : |
QY      355 -----TKTDNNNSNFTWTGASKY--NLNGRESIIINPGT---AMASHKDEDDEKFFMS 401
        : ||| | : | : | : | : | : | : | : | : | : |
Db      496 AAGRGAQTDENQAA--DGDPRYAFGRHQGOKTTTGETPEREFTYIAHQDT----- 544
        : ||| | : | : | : | : | : | : | : | : | : |
QY      402 GWMIPEGESAGASNTALD-NVMITDEEEIKATNPVATERFGTVAVNFQSSSTDPATGDVH 460
        : ||::: : : : : ||| | : | : | : | : | : | : |
Db      545 -----GRYPAGDWIQININFNPVTINDVLLPTDPIG---GKTGINY--TNIFNTYGPLT 593
        : ||::: : : : : ||| | : | : | : | : | : | : |
QY      461 AMGALFGMWVQDRDVYLQGPWAKIPHDTGHFHSPLMGGFLK-----NPP 507
        | : | : | | | | | | | | | | | | | | | | |
Db      594 ALNNVP-----PVYPNGQIMDKEDFTD-----LKPRLHVNAFVCQNNCP 633
        : ||::: : : : : ||| | : | : | : | : | : | : |
QY      508 PQLIKNTVPVAN---PPAEFSATKFASFITQYSTGOVSEIEMELQKENSKRWNDEVQY 564
        : ||::: : : : : ||| | : | : | : | : | : | : |
Db      634 GOLFYVAAPNLITNEYDPDASANMSR---ITYYSDFWWKKGKLVPKAKLRASHTWNP IQM 689
        : ||::: : : : : ||| | : | : | : | : | : | : |
QY      565 TSNYAKSANVDFTVDNGLYTEPRPIG 591
        : ||| | | | | | | | | | | | | | | |
Db      690 SIN-----VDNQFNYL-PNNIG 705
        : ||| | | | | | | | | | | | | | | |

```

```

RESULT 14
COAT_FPV19
ID COAT_FPV19 STANDARD; PRT; 727 AA.
AC P24840;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Coat protein VP1 [Contains: Coat protein VP2].
OS Feline panleukopenia virus (strain 193) (FPV).
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirus.
OX NCBI_TaxID=10787;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=193/70;
RX MEDLINE=91073139; PubMed=2174965;
RA Martyn J.C., Davidson B.E., Studdert M.J.;
RT "Nucleotide sequence of feline panleukopenia virus: comparison with
RL canine parvovirus identifies host-specific differences.";
RN J. Gen. Virol. 71:2747-2753(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CU-4;
RX MEDLINE=91272479; PubMed=1647068;
RA Parrish C.R.;
RT "Mapping specific functions in the capsid structure of canine
RL parvovirus and feline panleukopenia virus using infectious plasmid
RT clones.";
RN Virology 183:195-205(1991).
CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC	EMBL; X55115; CAA38911.1; -	727	AA; 80386	MM; 648596C09B621FF5	CRC64; 648596C09B621FF5
DR	EMBL; M38246; AAC37928.1; -	144	727	COAT PROTEIN VP1.	(POTENTIAL).
DR	EMBL; M38246; AAC37929.1; -	165	190	COAT PROTEIN VP2.	(POTENTIAL).
DR	PIR; B36608; VCPVFP.	168	190	GLY-RICH.	(POTENTIAL).
DR	HSSP; P30129; 4DPV.	190	190	N-LINKED (GLCNAC. . .)	(POTENTIAL).
DR	InterPro; IPR001403; Parvo_coat.	190	190	N-LINKED (GLCNAC. . .)	(POTENTIAL).
DR	Pfam; PF00740; Parvo_coat; 1.	190	190	N-LINKED (GLCNAC. . .)	(POTENTIAL).
KM	Coat protein; Glycoprotein.	190	190	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CHAIN 1 727	190	190	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CHAIN 144 727	190	190	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	DOMAIN 165 190	190	190	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD 168 168	190	190	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD 190 190	190	190	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD 207 207	190	190	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD 323 323	190	190	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD 586 586	190	190	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD 648 648	190	190	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD 660 660	190	190	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	DOMAIN 165 180	190	190	GLY-RICH.	(POTENTIAL).
FT	SEQUENCE 727 AA; 80386 MM; 648596C09B621FF5	190	190	CRC64; 648596C09B621FF5	

Query Match	6.4%;	Score 209;	DB 1;	Length 727;
Best Local Similarity	20.1%;	Pred. No. 1.7e-06;		
Matches 138;	Conservative 101;	Mismatches 260;	Indels 188;	Gaps 34;

QY	7	RVEQSPQEPDSSSGIGKTGQOPAKKRLNFGQTGDSSEVP----	DPQPLGEPPA--TPAA	60	
Db	110	KPTKRSKPPEPHIFINLAK-----	KKKAGAGQVKRDNLAFMSDGAQVDPGGQCAVNERA	163	
QY	61	VGPPTMASGGGAPMADNNEGADGVNASGNWH-----	CDSTWLGDRVITTTSTRTWALP	113	
Db	164	TGSGNGSGGG-----	GGSGGVGISTGTENNQTEFKLENGWV--EITANSRLVHLN	215	
QY	114	TYNNHLKYQI-----	SSASTGASNDNHYFGYSTPWGYFDENRFHCHFSPPDQRLINN	166	
Db	216	MESENYKRVVNNMDKTA VKGNMALLDHIHQIVTPMSLVDANAMGWFMFPGMQILVNT		275	
QY	167	NWGFRRPKRLNFKLENIQVEVT--	TNDGVTTIANNLTSTVQVFSDEYQLPYVLGSAHQ	223	
Db	276	MSELHLVSFEQEIFNVVLKTVSESATQPTKYVNNDLTASLMVALDSNNTMFTPAAMRS		335	
QY	224	GCL-----	PPFPA-----DVEMIPQGYGLTLNNGSAVGRS-----	SFYCL	259
Db	336	ETLGFYYPWKPTIPTPMWRYXFQWDRTLPSH-----	TGTSGPTNVYHGTDPDQVQYTTI		389
QY	260	E-YFPSQMLRTGNMF-TFSYTFEEVPEHSSYAHQSQSLDRLMN--	PLIDQYLYLNRTQNO		315
Db	390	ENSVPVHLRLRTGDEFATGTFPFCKP--	CRLTHWTQTRALGLP-----	FLNSLPQS	440
QY	316	SGSAQNKDLLFSRGSFAGMS-----	VQPKN-WLPGPCYRQQRVSK-----		354
Db	441	EGATNFGDIGVQDKRGRVTOQNGNTDYITEATIMRPAEVGYSAPYYSFEASTGPEKPTPI			500
QY	355	-----	TKTDNNSNFTWTGASKY---NLNGRESLIINPGT---AMASHKDEDEKFFPMS		401
Db	501	AAGRGAQTDENQAA--	DGDPRTAFGRQHGQKTTTGETPERFTYIAHQDT-----		549
QY	402	GVMIFGKESAGASNTALD-NVMTDEEBEIKATNPVATERFGTVAVNFOSSSTDPATGDVH			460
Db	550	-----	GRYPEGDWIONINENLPTVNDVLLPTDPIG---GKTGINY--TNINFTYGGPLT		598
QY	461	AMGALLPGMWVQDRDYYLQGPIMAKIPHTDGHFHPSPLMGGFGLK-----	NPP		507
Db	599	ALNNVP-----	PVYPNGQIWDKEFDTD-----LKPRLHVNAFPVCQNNCP		638
QY	508	POILLIKNTPVPAN--	PRAEFSATKFASEFITQYSTGQVSEIEMWLOKENSKEWNEBEVQY		564



Db 639 GOLFKVADNULJNEYDPDASANMSR-----ITYSDPFWMKGKLVFKAKLRASHWTWNPICQM 694  
QY 565 TSNTAKSANVDFTVNNGLYTEPRPIG 591  
Db 695 SIN-----VDNQFNIV-PNNIG 710

RESULT 15

COAT\_PAVCB STANDARD; PRT; 727 AA.  
ID COAT\_PAVCB  
AC Q11213;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Coat protein VP1 (Contains: Coat protein VP2).  
OS Canine parvovirus (strain B) (CPV).  
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.  
OX NCBI\_TaxID=59284;

RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91272479; PubMed=1647068;  
RA Parrish C.R.;  
RT "Mapping specific functions in the capsid structure of canine parvovirus and feline panleukopenia virus using infectious plasmid clones."  
RL Virology 183:195-205(1991).  
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS, OF A COMBINATION OF VP2, VP3, AND SOME VP1.  
CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; M38245; AAB02799.1; -.  
DR EMBL; M38245; AAB02800.1; -.  
DR HSSP; P30129; ADPV.  
DR InterPro; IPR001403; Parvo\_coat.  
DR Pfam; PF00740; Parvo\_coat; 1.  
KW Coat protein; Glycoprotein.

FT CHAIN 1 727 COAT PROTEIN VP1.  
FT CHAIN 144 727 COAT PROTEIN VP2.  
FT CARBOHYD 168 168 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 190 190 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 207 207 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 323 323 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 586 586 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 648 648 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 660 660 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 727 AA; 80342 MW; 115F3E1A79098EBE CRC64;

Query Match 6.4%; Score 208; DB 1; Length 727;  
Best Local Similarity 19.8%; Pred. No. 2e-06;  
Matches 137; Conservative 96; Mismatches 260; Indels 200; Gaps 31;

QY 7 RPVEQSPQEPDSSSGIGKTGQPAKKRLNFGQTGDSSEVP---DPQPLGEPPA--TPAA 60  
Db 110 KPTKRSKPPPHIFINLAK-----KKKAGAGQVNRDNLAPMSDGAIVQPDGQPAVRNERA 163  
QY 61 VGPTMASGGGAPMADNNEGADVNASGNWH-----CDSTWLGDRAVITSTRTWALP 113  
Db 164 TSGSGSGGGG-----GGSGGSGVISTGTFTNNQTEFKFLENGWV--EITANSSRLVHLN 215  
QY 114 TYNHLYKQI-----SSASTGASNDNHYFGYSTPWGYFDENRPHCHFSPRDWQRLINN 166  
Db 216 MPESSENYRNVVNNMDKTAVNGMALDDIHAQIVTPWSLVDANAMGVWFNPGDWQLIVNT 275  
QY 167 NWGFRPKRLNFKLFIQVKEVT--TNDGVTTIANNLSTVQVPSDSEYQLPYVLGSAHQ 223

Db 276 MSELHLVSFEQEIFNVVLKTVSESATQPTKVYNNDLTASLWALDSNNTMPTPAMRS 335  
QY 224 GCL-----PPFPA-----DVEMIPQYGLTLNNGSAVGRS-----SFYCL 259  
Db 336 ETLCGFYFWKPTIPTFWRYYFQMDRTLIPSH-----TGTSGTPTNIYHGTDPDDVQFYTI 389  
QY 260 E-YFPSQMLRTGNF-TFSYTFEEVPEFHSYASHQSILRLMN--PLIDQYLYLNRTONQ 315  
Db 390 ENSVAVHLRLTGDEFATGTFEFDCKP--CRLTHTWQTNRALGLP-----FLNSLPQS 440  
QY 316 SGSAQNKDLLFSRGSFAGMSVQPKNWLPGPCYRQQRVSKTKTDNNNSNF-----TW 366  
Db 441 EGATNMGDI-----GV-----QDQKRGVTQMGTNTYITEATIMRPAE 478  
QY 367 TGASKYNLNGRESIINP-----GTAMASHKDEDEKFPMSGVMIFGKESAGASNTALD-- 419  
Db 479 VGYSAPFYYSFEASTQGFPKPTIAAGRGGAQTDENQADGNPRYAFGRQHGQKTTTGETP 538  
QY 420 -----NMITDEEIKATNPVATERFGTVAVNFQSSSTD 454  
Db 539 ERTYIAHQDTGRYPREGDWIONINENLPVTINDNVLPLTDPIG---GKTGINY--TNIFN 592  
QY 455 ATGDVHAMGALPGMWQDRDYLLQGPPIWAKIPHTDGHHPSPLMGGFGLK----- 504  
Db 593 TYGPPLTALNVP-----PYVNGQIWDKEFDTD-----LKRPLHVNAPEV 632  
QY 505 --NPPQILIKNTPVPAN---PFAFSATKFASTLOYSTGQVSVEIEWELOKNSRW 558  
Db 633 CONNCPQGLFKVAPNLJNEYDPDASANMSR-----ITYSDPFWMKGKLVFKAKLRASHWTW 688  
QY 559 NPEVQYTSNYAKSANVDFTVNNGLYTEPRPIG 591  
Db 689 NPIQQMISIN-----VDNQFNIV-PSNIG 710

Search completed: July 17, 2003, 18:32:05  
Job time : 14.8582 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 2003, 18:26:09 ; Search time 44.5484 Seconds  
(without alignments)  
2770.518 Million cell updates/sec

Title: US-09-807-802A-15  
Perfect score: 3251  
Sequence: 1 TAPGKKRPVEQSPQEPDSSS.....NNGLYTEPRPIGTRYLTRPL 599

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : SPTREMBL 21:\*
- 1: sp\_archaea:\*
  - 2: sp\_bacteria:\*
  - 3: sp\_fungi:\*
  - 4: sp\_human:\*
  - 5: sp\_invertebrate:\*
  - 6: sp\_mammal:\*
  - 7: sp\_mhc:\*
  - 8: sp\_organelle:\*
  - 9: sp\_phase:\*
  - 10: sp\_plant:\*
  - 11: sp\_rodent:\*
  - 12: sp\_virus:\*
  - 13: sp Vertebrate:\*
  - 14: sp\_unclassified:\*
  - 15: sp\_rvirus:\*
  - 16: sp\_bacteriaph:\*
  - 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3251	100.0	736	12	Q9WBP8	Q9wbp8 adeno-associ
2	3229	99.3	736	12	O56137	O56137 adeno-associ
3	2832	87.1	736	12	O56139	O56139 adeno-associ
4	2815	86.6	736	12	Q65311	Q65311 adeno-associ
5	2764.5	85.0	735	12	O56652	O56652 adeno-associ
6	2759.5	84.9	598	12	O56653	O56653 adeno-associ
7	2481.5	76.3	533	12	O92917	O92917 adeno-associ
8	1830.5	56.3	734	12	O41855	O41855 adeno-associ
9	1717	52.8	732	12	O67666	O67666 goose parvo
10	1714	52.7	732	12	O83290	O83290 muscovy duc
11	1712	52.7	587	12	Q67667	Q67667 goose parvo
12	1712	52.7	732	12	Q8V395	Q8v395 goose parvo
13	1702	52.4	732	12	Q65444	Q65444 barbarie du
14	1701	52.3	732	12	Q83289	Q83289 muscovy duc
15	1700.5	52.3	724	12	Q9Y1J1	Q9y1j1 adeno-associ
16	1698	52.2	534	12	Q67668	Q67668 goose parvo

17	1695	52.1	587	12	Q65445	Q65445 barbarie du
18	1676	51.6	534	12	Q65446	Q65446 barbarie du
19	1540	47.4	676	12	Q67672	Q67672 goose parvo
20	715	22.0	179	12	Q9WA24	Q9wa24 goose parvo
21	715	22.0	179	12	Q9W8U7	Q9w8u7 duck parvo
22	711	21.9	179	12	Q9WN18	Q9wn18 duck parvo
23	709	21.8	179	12	Q9WN19	Q9wn19 duck parvo
24	707	21.7	179	12	Q9WN20	Q9wn20 goose parvo
25	632.5	19.5	703	12	Q8QQV5	Q8qqv5 minute viru
26	618	19.0	571	12	Q8QQV4	Q8qqv4 minute viru
27	545.5	16.8	947	12	Q918U9	Q918u9 bovine parv
28	508	15.6	781	12	Q9PZT0	Q9pzt0 human parvo
29	501	15.4	554	12	Q9PZS9	Q9pzs9 human parvo
30	501	15.4	781	12	Q9JGP8	Q9jgp8 human parvo
31	500.5	15.4	773	12	Q913X1	Q913x1 human parvo
32	499.5	15.4	785	12	Q9J0X4	Q9j0x4 pig-tailed
33	499	15.3	781	12	P89318	P89318 human parvo
34	499	15.3	781	12	P89319	P89319 human parvo
35	498.5	15.3	781	12	P89317	P89317 human parvo
36	497.5	15.3	781	12	Q912B8	Q912b8 human eryth
37	497	15.3	769	12	Q9PZT4	Q9pzt4 human parvo
38	497	15.3	781	12	Q9JGS0	Q9jgs0 human parvo
39	497	15.3	781	12	P90221	P90221 human parvo
40	497	15.3	781	12	P90222	P90222 human parvo
41	497	15.3	781	12	P90223	P90223 human parvo
42	497	15.3	781	12	P90224	P90224 human parvo
43	497	15.3	781	12	Q85191	Q85191 human parvo
44	497	15.3	781	12	Q85117	Q85117 human parvo
45	497	15.3	781	12	Q90200	Q90200 human parvo

ALIGNMENTS

RESULT 1									
ID	Q9WBP8	PRELIMINARY;	PRT;	736	AA.				
AC	Q9WBP8;								
DT	01-NOV-1999 (TREMBLrel. 12, Created)								
DT	01-NOV-1999 (TREMBLrel. 12, Last sequence update)								
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)								
DE	Capsid protein.								
OS	adeno-associated virus 1.								
OC	Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.								
OX	NCBI_TaxID=85106;								
RN	[1]								
RP	SEQUENCE FROM N.A.								
RX	MEDLINE=99214338; PubMed=10196295;								
RA	Xiao W., Chirmule N., Berta S.C., McCullough B., Gao G., Wilson J.M.;								
RT	"Gene therapy vectors based on adeno-associated virus type 1.";								
RL	J. Virol. 73:3994-4003(1999).								
RN	[2]								
RP	SEQUENCE FROM N.A.								
RA	Xiao W., Wilson J.M.;								
RL	Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.								
DR	EMBL; AF063497; AAD27757.1; -.								
DR	InterPro; IPR001403; Parvo_coat.								
DR	Pfam; PF00740; Parvo_coat; 1.								
DR	SEQUENCE 736 AA; 81375 MW; CFAFB9BD5CD0595 CRC64;								
Query Match									
Best Local Similarity		100.0%;	Score 3251;	DB 12;	Length 736;				
Matches 599;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;				
QY	1	TAPGKKRPVEQSPQEPDSSSGIGKTGQOPAKKRLNFGQTGDSVSPDPQPLGEPATPAA	60						
DB	138	TAPGKKRPVEQSPQEPDSSSGIGKTGQOPAKKRLNFGQTGDSVSPDPQPLGEPATPAA	197						
QY	61	VGPTTASGGGAPMADNNEGADGVNAGSNWHCDSTWLGDRVITTTSTRTWALPTYNNHLY	120						
DB	198	VGPTTASGGGAPMADNNEGADGVNAGSNWHCDSTWLGDRVITTTSTRTWALPTYNNHLY	257						
QY	121	KQISSASTGASNDNHYFGYSTPWGYPDENRFHCHFSPPDWQRLINNMGFRPKRLNFKLF	180						



```
|||||
Db 258 KOISSASTGASNDNHFGYSTPWGYFDENRFHCHFSPRDQRLINNNGFRPKRLNFKLF 317
QY 181 NIQVEVTNDGVTTIANNLTSTVQVFSDEYQLPYVLGSAHQCLPPFPADVEMIPQYG 240
Db 318 NIQVEVTNDGVTTIANNLTSTVQVFSDEYQLPYVLGSAHQCLPPFPADVEMIPQYG 377
QY 241 YLTNNGSQAVGRSSFYCLEYFSPQMLRTGNFTFSYTFEEVPHSSSYAHSQSLDRLMNP 300
Db 378 YLTNNGSQAVGRSSFYCLEYFSPQMLRTGNFTFSYTFEEVPHSSSYAHSQSLDRLMNP 437
QY 301 LIDQLYLNRTONQSGSAQNKDLLFSRGSFAGMSVQPKNWLPGPCYRQQRVSKTKTDNN 360
Db 438 LIDQLYLNRTONQSGSAQNKDLLFSRGSFAGMSVQPKNWLPGPCYRQQRVSKTKTDNN 497
QY 361 NSNFTWTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMGVMIFGKESAGASNTALDN 420
Db 498 NSNFTWTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMGVMIFGKESAGASNTALDN 557
QY 421 VMITDEEIKATNPVATERFGTVAVNFQSSSTDPAATGDVHAMGALPGMWQDRDVLQGP 480
Db 558 VMITDEEIKATNPVATERFGTVAVNFQSSSTDPAATGDVHAMGALPGMWQDRDVLQGP 617
QY 481 IWAKIPHTDGHFHPSPPLMGFGGLKNPPQILIKNTVPANPAPAEFSATKFAFITQYSTG 540
Db 618 IWAKIPHTDGHFHPSPPLMGFGGLKNPPQILIKNTVPANPAPAEFSATKFAFITQYSTG 677
QY 541 QVSVEIEMELQKENSKRWNPEVOYTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 599
Db 678 QVSVEIEMELQKENSKRWNPEVOYTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 736
```

RESULT 2

```
056137 PRELIMINARY; PRT; 736 AA.
ID 056137
AC 056137;
DT 01-JUN-1998 (TReMBLrel. 06, Created)
DT 01-JUN-1998 (TReMBLrel. 06, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE Capsid protein VP1.
OS adeno-associated virus 6.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
OX NCBI_TaxID=68558;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98080418; PubMed=9420229;
RA Rutledge E.A., Halbert C.L., Russell D.W.;
RT "Infectious clones and vectors derived from adeno-associated virus (AAV) serotypes other than AAV type 2.";
RL J. Virol. 72:309-319(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Rutledge E.A., Russell D.W.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF028704; AAB95450.1;
DR InterPro; IPR001403; Parvo_coat.
DR Pfam; PF00740; Parvo_coat; 1.
SQ SEQUENCE 736 AA; 81411 MW; 311217A089C565F5 CRC64;

Query Match 99.3%; Score 3229; DB 12; Length 736;
Best Local Similarity 99.2%; Pred. No. 3.6e-223;
Matches 594; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 TAPGKRPVEQSPQEPDSSSGIGTGOQPAKRLNFGQTGDSSEVPDPQPLGEPATPAA 60
Db 138 TAPGKRPVEQSPQEPDSSSGIGTGOQPAKRLNFGQTGDSSEVPDPQPLGEPATPAA 197
QY 61 VGPTTASGGGAPMADNNEGADGVNASGNWCHDSTWLGDRVITTTSTRTWALPTYNNHLY 120
Db 198 VGPTTASGGGAPMADNNEGADGVNASGNWCHDSTWLGDRVITTTSTRTWALPTYNNHLY 257
QY 121 KOISSASTGASNDNHFGYSTPWGYFDENRFHCHFSPRDQRLINNNGFRPKRLNFKLF 180
Db 258 KOISSASTGASNDNHFGYSTPWGYFDENRFHCHFSPRDQRLINNNGFRPKRLNFKLF 316
```

```
Db 258 KOISSASTGASNDNHFGYSTPWGYFDENRFHCHFSPRDQRLINNNGFRPKRLNFKLF 317
QY 181 NIQVEVTNDGVTTIANNLTSTVQVFSDEYQLPYVLGSAHQCLPPFPADVEMIPQYG 240
Db 318 NIQVEVTNDGVTTIANNLTSTVQVFSDEYQLPYVLGSAHQCLPPFPADVEMIPQYG 377
QY 241 YLTNNGSQAVGRSSFYCLEYFSPQMLRTGNFTFSYTFEEVPHSSSYAHSQSLDRLMNP 300
Db 378 YLTNNGSQAVGRSSFYCLEYFSPQMLRTGNFTFSYTFEEVPHSSSYAHSQSLDRLMNP 437
QY 301 LIDQLYLNRTONQSGSAQNKDLLFSRGSFAGMSVQPKNWLPGPCYRQQRVSKTKTDNN 360
Db 438 LIDQLYLNRTONQSGSAQNKDLLFSRGSFAGMSVQPKNWLPGPCYRQQRVSKTKTDNN 497
QY 361 NSNFTWTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMGVMIFGKESAGASNTALDN 420
Db 498 NSNFTWTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMGVMIFGKESAGASNTALDN 557
QY 421 VMITDEEIKATNPVATERFGTVAVNFQSSSTDPAATGDVHAMGALPGMWQDRDVLQGP 480
Db 558 VMITDEEIKATNPVATERFGTVAVNFQSSSTDPAATGDVHAMGALPGMWQDRDVLQGP 617
QY 481 IWAKIPHTDGHFHPSPPLMGFGGLKNPPQILIKNTVPANPAPAEFSATKFAFITQYSTG 540
Db 618 IWAKIPHTDGHFHPSPPLMGFGGLKNPPQILIKNTVPANPAPAEFSATKFAFITQYSTG 677
QY 541 QVSVEIEMELQKENSKRWNPEVOYTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 599
Db 678 QVSVEIEMELQKENSKRWNPEVOYTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 736
```

RESULT 3

```
056139 PRELIMINARY; PRT; 736 AA.
ID 056139
AC 056139;
DT 01-JUN-1998 (TReMBLrel. 06, Created)
DT 01-JUN-1998 (TReMBLrel. 06, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE Capsid protein VP1.
OS adeno-associated virus 3B.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
OX NCBI_TaxID=68742;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98080418; PubMed=9420229;
RA Rutledge E.A., Halbert C.L., Russell D.W.;
RT "Infectious clones and vectors derived from adeno-associated virus (AAV) serotypes other than AAV type 2.";
RL J. Virol. 72:309-319(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Rutledge E.A., Russell D.W.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF028705; AAB95452.1;
DR InterPro; IPR001403; Parvo_coat.
DR Pfam; PF00740; Parvo_coat; 1.
SQ SEQUENCE 736 AA; 81906 MW; DD52331AD5F0D70F CRC64;

Query Match 87.1%; Score 2832; DB 12; Length 736;
Best Local Similarity 85.7%; Pred. No. 1.1e-194;
Matches 514; Conservative 34; Mismatches 50; Indels 2; Gaps 2;

QY 1 TAPGKRPVEQSPQEPDSSSGIGTGOQPAKRLNFGQTGDSSEVPDPQPLGEPATPAA 60
Db 138 TAPGKRPVDQSPQEPDSSSGVGSGKQPAKRLNFGQTGDSSEVPDPQPLGEPAPPTS 197
QY 61 VGPTTASGGGAPMADNNEGADGVNASGNWCHDSTWLGDRVITTTSTRTWALPTYNNHLY 120
Db 198 LGSNTMASGGGAPMADNNEGADGVNASGNWCHDSQWLGDRVITTTSTRTWALPTYNNHLY 257
QY 121 KOISSASTGASNDNHFGYSTPWGYFDENRFHCHFSPRDQRLINNNGFRPKRLNFKLF 180
Db 258 KOISSQS-GASNDNHFGYSTPWGYFDENRFHCHFSPRDQRLINNNGFRPKLSFKLF 316
```

QY 181 NIQKEVTNDGVTTIANNLTSTVQVFSDEYQLPYVLGSAHQGLPPFPADVEMIPQYG 240  
DB 317 NIQKEVTQNDGTTTIANNLSTVQVFTSEYQLPYVLGSAHQGLPPFPADVEMIPQYG 376  
QY 241 YLTINNGSQAVGRSSFYCLEYFPSPQMLRTGNFTFSYTFEEVPEHSSYAHQSQSLDRLMP 300  
DB 377 YLTINNGSQAVGRSSFYCLEYFPSPQMLRTGNFTFSYTFEEVPEHSSYAHQSQSLDRLMP 436  
QY 301 LIDQYLYLNRTQ-NQSGSAQNKDLLFSRGSPPAGMSVQPKNWLPGPCYRQORVSKTKTDN 359  
DB 437 LIDQYLYLNRTQGTSTGTTNQSRLIFSQAGPQSMISQARNWLPGPCYRQORLSKTANDN 496  
QY 360 NNSNFTWTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMSGVMIFGKESAGASNTALD 419  
DB 497 NNSNFTWTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMSGVMIFGKESAGASNTALD 556  
QY 420 NVMITDEEIKATNPVATERFGTVAVNFQSSSTDPAATGVDHAMGALPGMWQDRDVLQ 479  
DB 557 NVMITDEEIKATNPVATERFGTVAVNFQSSSTDPAATGVDHAMGALPGMWQDRDVLQ 616  
QY 480 PIMAKIPHTDGHFHPSPPLMGFGGLKNPPQILIKNTPVANPAPAFSATKFAFITQYST 539  
DB 617 PIMAKIPHTDGHFHPSPPLMGFGGLKNPPQILIKNTPVANPAPAFSATKFAFITQYST 676  
QY 540 GQVSVEIEMELQKENSKRWNPEVOYTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 599  
DB 677 GQVSVEIEMELQKENSKRWNPEVOYTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 736

RESULT 4

Q65311 PRELIMINARY; PRT; 736 AA.  
ID Q65311  
AC Q65311;  
DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
DE Capsid protein.  
OS adeno-associated virus 3.  
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.  
OX NCBI\_TaxID=46350;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=3H;  
RX MEDLINE=96266430; Pubmed=8661429;  
RA Muramatsu S., Mizukami H., Young N.S., Brown K.E.;  
RT "Nucleotide sequencing and generation of an infectious clone of adeno-associated virus 3."  
RT associated virus 3."  
RL virology 221:208-217(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=3H;  
RA Muramatsu S., Brown K.E.;  
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U48704; AAC55049.1; -.  
DR InterPro; IPR001403; Parvo\_coat.  
DR Pfam; PF00740; Parvo\_coat; 1.  
SQ SEQUENCE 736 AA; 81660 MW; AFFIEF47B5C67A10 CRC64;

Query Match 86.6%; Score 2815; DB 12; Length 736;  
Best Local Similarity 85.2%; Pred. No. 1.8e-193;  
Matches 511; Conservative 35; Mismatches 52; Indels 2; Gaps 2;

QY 1 TAPGKKRPVEQSPQEPDSSSGIGKTGQOPAKKRLNFGQTGDSSEVPDQPLGEPATPAA 60  
DB 138 TAPGKKGAVDQSPQEPDSSSGVGKSGKQAPARKRLNFGQTGDSSEVPDQPLGEPAPATS 197  
QY 61 VGPTMASGGAPMADNNEGADGVGNASGNWHDSTWLGDRTTSTRTALPTYNNHLY 120  
DB 198 LGSNTMASGGAPMADNNEGADGVGNSSGNWHDSTWLGDRTTSTRTALPTYNNHLY 257  
QY 121 KQISSASTGASNDNHYFGYSTPWGYFDNFRFHCHFSPRDWQRLINNNGFRPKRLNFKLF 180

DB 258 KQISSQS-GASNDNHYFGYSTPWGYFDNFRFHCHFSPRDWQRLINNNGFRPKRLNFKLF 316  
QY 181 NIQKEVTNDGVTTIANNLTSTVQVFSDEYQLPYVLGSAHQGLPPFPADVEMIPQYG 240  
DB 317 NIQKEVTQNDGTTTIANNLSTVQVFTSEYQLPYVLGSAHQGLPPFPADVEMIPQYG 376  
QY 241 YLTINNGSQAVGRSSFYCLEYFPSPQMLRTGNFTFSYTFEEVPEHSSYAHQSQSLDRLMP 300  
DB 377 YLTINNGSQAVGRSSFYCLEYFPSPQMLRTGNFTFSYTFEEVPEHSSYAHQSQSLDRLMP 436  
QY 301 LIDQYLYLNRTQ-NQSGSAQNKDLLFSRGSPPAGMSVQPKNWLPGPCYRQORVSKTKTDN 359  
DB 437 LIDQYLYLNRTQGTSTGTTNQSRLIFSQAGPQSMISQARNWLPGPCYRQORLSKTANDN 496  
QY 360 NNSNFTWTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMSGVMIFGKESAGASNTALD 419  
DB 497 NNSNFTWTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMSGVMIFGKESAGASNTALD 556  
QY 420 NVMITDEEIKATNPVATERFGTVAVNFQSSSTDPAATGVDHAMGALPGMWQDRDVLQ 479  
DB 557 NVMITDEEIKATNPVATERFGTVAVNFQSSSTDPAATGVDHAMGALPGMWQDRDVLQ 616  
QY 480 PIMAKIPHTDGHFHPSPPLMGFGGLKNPPQILIKNTPVANPAPAFSATKFAFITQYST 539  
DB 617 PIMAKIPHTDGHFHPSPPLMGFGGLKNPPQILIKNTPVANPAPAFSATKFAFITQYST 676  
QY 540 GQVSVEIEMELQKENSKRWNPEVOYTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 599  
DB 677 GQVSVEIEMELQKENSKRWNPEVOYTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 736

RESULT 5

O56652 PRELIMINARY; PRT; 735 AA.  
ID O56652  
AC O56652;  
DT 01-JUN-1998 (Tremblrel. 06, Created)  
DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE Major coat protein VP1.  
OS Adeno-associated virus 2 (AAV2).  
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.  
OX NCBI\_TaxID=10804;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC MEDLINE=95088582; Pubmed=7996133;  
RA Ruffing M., Heid H., Kleinschmidt J.A.;  
RT "Mutations in the carboxy terminus of adeno-associated virus 2 capsid proteins affect viral infectivity: lack of an RGD integrin-binding motif."  
RT J. Gen. Virol. 75:0-0(0).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Berns K.I., Bohenzky R.A., Cassinotti P., Colvin D., Donahue B.A.,  
RA Dull T., Hoyer M., Kleinschmidt J.A., Ruffing M., Snyder R.O.,  
RA Tratschin J.-D., Weitz M.;  
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF043303; AAC03780.1; -.  
DR InterPro; IPR001403; Parvo\_coat.  
DR Pfam; PF00740; Parvo\_coat; 1.  
KW Coat protein.  
FT VARIANT 76 D -> V.  
FT VARIANT 553 D -> N.  
FT VARIANT 567 T -> N.  
FT VARIANT 677 QV -> HV.  
FT VARIANT 710 V -> R.  
SQ SEQUENCE 735 AA; 81944 MW; 980BEEF46908390B CRC64;

Query Match 85.0%; Score 2764.5; DB 12; Length 735;  
Best Local Similarity 82.8%; Pred. No. 7.3e-190;  
Matches 496; Conservative 43; Mismatches 59; Indels 1; Gaps 1;

QY 1 TAPGKKRPVEQSPQEPDSSSGIGKTGQOPAKKRLNFGQTGDSSEVPDQPLGEPATPAA 60

Db 138 TAPGKKRPVEHSPVEPDDSSSGTGAKGQOPARKRLNFGQTGDADSVDPDQPLGPPAPASG 197  
QY 61 VGPPTMASGGAPMADNNEGADGVGNASGNWHDSTWLGDRIITTTSTRTWALPTYNNHLY 120  
Db 198 LGTNTMATGSGAPMADNNEGADGVGNSSGNWHCDSTWMDRIVITTTSTRTWALPTYNNHLY 257  
QY 121 KOISSASTGASNDNHFGYSTPWGCFDENFRFHCHFSPRDWORLINNNGFRPKRLNFKLF 180  
Db 258 KOISSQS-GASNDNHFGYSTPWGCFDENFRFHCHFSPRDWORLINNNGFRPKRLNFKLF 316  
QY 181 NIQVEVTNDGVTTIANNLTSTVOVPSDSEYQLPYVLGSAHQGLPPFPADVFMIPQY 240  
Db 317 NIQVEVTNDGVTTIANNLTSTVOVPSDSEYQLPYVLGSAHQGLPPFPADVFMIPQY 376  
QY 241 YLTLMNGSAVGRSSFYCLEYFPSPQMLRTGNFTFSYTFEEVPHSSYAHQSOLDRLMP 300  
Db 377 YLTLMNGSAVGRSSFYCLEYFPSPQMLRTGNFTFSYTFEEVPHSSYAHQSOLDRLMP 436  
QY 301 LIDQYLYLNRTONOGSGAQNKDLFSRGSFAGMSVQPKNMLPGPCYRQORVSKTKTDNN 360  
Db 437 LIDQYLYLSRTNTPSGTTTQSLRQFSQAGASDIRDQSRNMLPGPCYRQORVSKTSADNN 496  
QY 361 NSNFTWTGASKYNLNGRESINPGTAMASHKDEDEKFFPMGVMIFGKESAGASNTALDN 420  
Db 497 NSEYSMTGATKYHLNGBDSLVPNGPAMASHKDEEKFPPQSGVLI FGKQSEKTNVDIBK 556  
QY 421 VMITDEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMGALPGMWQDRDVLQGP 480  
Db 557 VMITDEEIRTNTPVATEQYGSVSTNLQRGNRQAATADVNTQGVLPGMWQDRDVLQGP 616  
QY 481 IWAKIPHTDGHFHPSPLMGGFGLKNPPQILIKNTVPANPAPAFSATKFAFITQYSTG 540  
Db 617 IWAKIPHTDGHFHPSPLMGGFGLKNPPQILIKNTVPANPAPAFSATKFAFITQYSTG 676  
QY 541 QVSVEIEMELQENSKRMNPEVOYTSNYAKSANDFTVDNGLYTEPRPIGTRYLTRPL 599  
Db 677 QVSVEIEMELQENSKRMNPEIQTYSNYSKSVNDFTVDNGLYTEPRPIGTRYLTRNL 735

RESULT 6

056653 PRELIMINARY; PRT; 598 AA.  
ID 056653;  
AC 056653;  
DT 01-JUN-1998 (Tremblrel. 06, Created)  
DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE Major coat protein VP2.  
OS Adeno-associated virus 2 (AAV2).  
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.  
OX NCBI\_TaxID=10804;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95088582; Pubmed=7996133;  
RA Ruffing M., Heid H., Kleinschmidt J.A.;  
RT "Mutations in the carboxy terminus of adeno-associated virus 2 capsid  
RT proteins affect viral infectivity: lack of an RGD integrin-binding  
RT motif.";  
RT J. Gen. Virol. 75:0-0(0).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX Berns K.I., Bohenzky R.A., Cassinotti P., Colvin D., Donahue B.A.,  
RA Dull T., Horer M., Kleinschmidt J.A., Ruffing M., Snyder R.O.,  
RA Tratschin J.-D., Weitz M.;  
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF043303; AAC03778.1; -;  
DR InterPro; IPR001403; Parvo\_coat.  
DR Pfam; PF00740; Parvo\_coat; 1.  
KW Coat protein.  
FT VARIANT 416 ' 416 D -> N.  
FT VARIANT 430 430 T -> N.  
FT VARIANT 540 541 QV -> HV.  
FT VARIANT 573 573 V -> R.  
SQ SEQUENCE 598 AA; 66619 MW; 070811ED9368E934 CRC64;

Query Match 84.9%; Score 2759.5; DB 12; Length 598;  
Best Local Similarity 82.8%; Pred. No. 1.3e-189;  
Matches 495; Conservative 43; Mismatches 59; Indels 1; Gaps 1;

QY 2 APGKKRPVEHSPVEPDDSSSGIGTKGQOPAKKRLNFGQTGDSESVDPDQPLGEPAPAPAV 61  
Db 2 APGKKRPVEHSPVEPDDSSSGTGAKGQOPARKRLNFGQTGDADSVDPDQPLGPPAPASGL 61  
QY 62 GPTTMASGGAPMADNNEGADGVGNASGNWHDSTWLGDRIITTTSTRTWALPTYNNHLYK 121  
Db 62 GTNTMATGSGAPMADNNEGADGVGNSSGNWHCDSTWMDRIVITTTSTRTWALPTYNNHLYK 121  
QY 122 QISSASTGASNDNHFGYSTPWGCFDENFRFHCHFSPRDWORLINNNGFRPKRLNFKLFN 181  
Db 122 QISSQS-GASNDNHFGYSTPWGCFDENFRFHCHFSPRDWORLINNNGFRPKRLNFKLFN 180  
QY 182 IOVKEVTNDGVTTIANNLTSTVOVPSDSEYQLPYVLGSAHQGLPPFPADVFMIPQY 241  
Db 181 IOVKEVTNDGVTTIANNLTSTVOVPSDSEYQLPYVLGSAHQGLPPFPADVFMIPQY 240  
QY 242 LTLNNGSAVGRSSFYCLEYFPSPQMLRTGNFTFSYTFEEVPHSSYAHQSOLDRLMP 301  
Db 241 LTLNNGSAVGRSSFYCLEYFPSPQMLRTGNFTFSYTFEEVPHSSYAHQSOLDRLMP 300  
QY 302 IDQYLYLNRTONOGSGAQNKDLFSRGSFAGMSVQPKNMLPGPCYRQORVSKTKTDNN 361  
Db 301 IDQYLYLSRTNTPSGTTTQSLRQFSQAGASDIRDQSRNMLPGPCYRQORVSKTSADNN 360  
QY 362 SNFTWTGASKYNLNGRESINPGTAMASHKDEDEKFFPMGVMIFGKESAGASNTALDNV 421  
Db 361 SEYSMTGATKYHLNGBDSLVPNGPAMASHKDEEKFPPQSGVLI FGKQSEKTNVDIEKV 420  
QY 422 MITDEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMGALPGMWQDRDVLQGP 481  
Db 421 MITDEEIRTNTPVATEQYGSVSTNLQRGNRQAATADVNTQGVLPGMWQDRDVLQGP 480  
QY 482 WAKIPHTDGHFHPSPLMGGFGLKNPPQILIKNTVPANPAPAFSATKFAFITQYSTGQ 541  
Db 481 WAKIPHTDGHFHPSPLMGGFGLKNPPQILIKNTVPANPAPAFSATKFAFITQYSTGQ 540  
QY 542 VSVSEIEMELQENSKRMNPEVOYTSNYAKSANDFTVDNGLYTEPRPIGTRYLTRPL 599  
Db 541 VSVSEIEMELQENSKRMNPEIQTYSNYSKSVNDFTVDNGLYTEPRPIGTRYLTRNL 598

RESULT 7

092917 PRELIMINARY; PRT; 533 AA.  
ID 092917;  
AC 092917;  
DT 01-NOV-1998 (Tremblrel. 08, Created)  
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)  
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
DE Major coat protein VP3.  
OS Adeno-associated virus 2 (AAV2).  
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.  
OX NCBI\_TaxID=10804;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95088582; Pubmed=7996133;  
RA Ruffing M., Heid H., Kleinschmidt J.A.;  
RT "Mutations in the carboxy terminus of adeno-associated virus 2 capsid  
RT proteins affect viral infectivity: lack of an RGD integrin-binding  
RT motif.";  
RT J. Gen. Virol. 75:0-0(0).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX Berns K.I., Bohenzky R.A., Cassinotti P., Colvin D., Donahue B.A.,  
RA Dull T., Horer M., Kleinschmidt J.A., Ruffing M., Snyder R.O.,  
RA Tratschin J.-D., Weitz M.;  
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF043303; AAC03779.1; -;  
DR InterPro; IPR001403; Parvo\_coat.



DR Pfam; PF00740; Parvo\_coat; 1.  
KM Coat protein.  
FT VARIANT 351 351 D -> N.  
FT VARIANT 365 365 T -> N.  
FT VARIANT 475 476 QV -> HV.  
FT VARIANT 508 508 V -> R.  
SQ SEQUENCE 533 AA; 60063 MM; 9E4D8BC25810D4F0 CRC64;

Query Match 76.3%; Score 2481.5; DB 12; Length 533;  
Best Local Similarity 83.3%; Pred. No. 9.2e-170;  
Matches 445; Conservative 37; Mismatches 51; Indels 1; Gaps 1;

QY 66 MASGGAPMADNNEGADGVGNASGNWCHDSTWLGDRVITSTRVMALEPTVNNHLYKQISS 125  
DB 1 MATGSGAPMADNNEGADGVGNSSGNWCHDSTWMDRVITSTRVMALEPTVNNHLYKQISS 60  
QY 126 ASTGASNDNHFGYSTPWGYPDFNRFHCHSPRDWQRLINNNWGFRPKRLNEKLFNIQVK 185  
DB 61 QS-GASNDNHFGYSTPWGYPDFNRFHCHSPRDWQRLINNNWGFRPKRLNEKLFNIQVK 119  
QY 186 EVTTNDGVTITANNLTSTVQVFSSEYQLPYVLGSAHQCLPPFPADVEMIPOYGYLTIN 245  
DB 120 EVTQNDGVTITANNLTSTVQVFTDSEYQLPYVLGSAHQCLPPFPADVEMVPOYGYLTIN 179  
QY 246 NGSQAVGRSSFYCLEYFSPSOMLRTGNNFTSYTFEEVFPFSSYAHQSGLDRMLNPLIDQY 305  
DB 180 NGSQAVGRSSFYCLEYFSPSOMLRTGNNFTSYTFEDVPFSSYAHQSGLDRMLNPLIDQY 239  
QY 306 LYYLNRTQNGSGSAQNKDLFSRGSFAGMSVQPKNWLPGPCYRQQRVSKTKTDNNNSNFT 365  
DB 240 LYYLSRTNTPSGTTTQSRLOFSQAGASDIRDQSRNWLPGPCYRQQRVSKTSADNNSEYS 299  
QY 366 WTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMGVMIFGKESAGASNTALDNVMTD 425  
DB 300 WTGATKYHLNGRDSLVPNGPAMASHKDEKFFPQSGVLI FGQGEKTNVDIEKVMITD 359  
QY 426 EEEIKATNPVATERFGTVAVNFQSSSTDPATGVDHAMGALPGMWODRDVYLQGPWAKI 485  
DB 360 EEEIRTNTPVATEQYGSVSTNLQGRNQAAATADVNTQGVLPGMWODRDVYLQGPWAKI 419  
QY 486 PHTDGHFHPSPLMGFGGLKNPPQILIKNTPVPANPPAEFSATKFAFITQYSTGQVSVE 545  
DB 420 PHTDGHFHPSPLMGFGGLKHPPIQILIKNTPVPANPSTTFSAAKFASFITQYSTGQVSVE 479  
QY 546 IEWELQKENS KRWNPVQYTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 599  
DB 480 IEWELQKENS KRWNPBEIQYTSNYKNSVNVDFTVDTNGVYSEPRPIGTRYLTRNL 533

## RESULT 8

Q41855 PRELIMINARY; PRT; 734 AA.  
AC 041855;  
DT 01-JAN-1998 (Tremblrel. 05, Created)  
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)  
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
DE Capsid.  
OS adeno-associated virus 4.  
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.  
OX NCBI\_TaxID=57579;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC VR-646;  
RX MEDLINE=97404695; PubMed=9261407;  
RA Chiorini J.A., Yang L., Liu Y., Safer B., Kotin R.M.;  
RT "Cloning of adeno-associated virus type 4 (AAV4) and generation of recombinant AAV4 particles.";  
RL J. Virol. 71:6823-6833(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC VR-646;  
RA Chiorini J.A., Yang L., Kotin R.M., Safer B.;  
RT Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL; U89790; AAC58045.1; -.  
DR InterPro; IPR001403; Parvo\_coat.  
DR Pfam; PF00740; Parvo\_coat; 1.  
SQ SEQUENCE 734 AA; 80639 MM; 616CC27A777BBE6F CRC64;

Query Match 56.3%; Score 1830.5; DB 12; Length 734;  
Best Local Similarity 57.8%; Pred. No. 7e-123;  
Matches 355; Conservative 74; Mismatches 154; Indels 31; Gaps 10;

QY 1 TAPGKKRPEVQSPQEPDSSSGIGTKGQPAKRLNF-GQTDSESVDPQPLGEPAATPA 59  
DB 137 TAPGKKRPLIESPQPDSSGTIGKKQPAKRLNFVEDETGAAGDPRPEGSTG-----A 190  
QY 60 AVGPTTMAAGGAPMADNNEGADGVGNASGNWCHDSTWLGDRVITSTRVMALEPTVNNHLY 119  
DB 191 MSDSEMRRAAAGAAVEGGQADGVGNASGDMHCDSTWSEGHVTTSTRTWLPTVNNHL 250  
QY 120 YKQISASTGASNDNHFGYSTPWGYPDFNRFHCHSPRDWQRLINNNWGFRPKRLNEKLF 179  
DB 251 YKRLGE---SLQSNTRYNGFSTPWGYPDFNRFHCHSPRDWQRLINNNWGFRPKARVYKI 306  
QY 180 FNIOYKEVTNDGVTITANNLTSTVQVFSSEYQLPYVLGSAHQCLPPFPADVEMIPOY 239  
DB 307 FNIOYKEVTNSGETTVANNLTSTVQIFADSSYELPYVMDAGQEGSLPPFPNDVFMVPOY 366  
QY 240 GY---LTNNGSQAVGRSSFYCLEYFSPSOMLRTGNNFTSYTFEEVFPFSSYAHQSGLDR 296  
DB 367 GYCGAVTGNTSQQQTDNNAFYCLEYFSPSOMLRTGNNFTSYTFEEVFPFSSYAHQSGLDR 426  
QY 297 LNMPLIDQYLYLNRTQN---QSGSAQNKDLFSRGSFAGMSVQPKNWLPGPCYRQQRV 352  
DB 427 LNMPLIDQYLMGLQSTTGTITNAGTATTN--FTKLRTNFSNFKKNWLPGPSIKQGF 483  
QY 353 SKTYDNNNSNFTWTGAS---KY---NLNGRESIINPGTAMASHKDEDEKFFPMGVMIF 405  
DB 484 SKTA--NQNKYIPATGSDSLIKETHSTLDGRWSALTGPMPMATAGPADSK-FSNSQLIF 540  
QY 406 FGKESAGASNTALDNVMTDEEIKATNPVATERFGTVAVNFQSSSTDPATGVDHAMGAL 465  
DB 541 AGPKQNGNTATVPGLTIFTSBEELATNATDITDMWGMLPGGDQNSNLTPTVDRLTALGAV 600  
QY 466 PGMWODRDVYLQGPWAKIPHTDGHFHPSPLMGFGGLKNPPQILIKNTPVPANPPAEF 525  
DB 601 PGMWQNRDIYYQGPWAKIPHTDGHFHPSPLIIGFGGLKHPPIQILIKNTPVPANPATTF 660  
QY 526 SATKFAFITQYSTGQVSVEIEWELQKENS KRWNPVQYTSNYAKSANVDFTVDNNGLYT 585  
DB 661 SSTPVNSFITQYSTGQVSQVQIDWEIQKERSKRWNPVQYTSNYGQONSLLWAPDAAGKKT 720  
QY 586 EPRPIGTRYLTRPL 599  
DB 721 EPRPIGTRYLTRPL 734

## RESULT 9

Q67666 PRELIMINARY; PRT; 732 AA.  
AC Q67666;  
DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE VP1.  
GN VP1.  
OS goose parvovirus.  
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.  
OX NCBI\_TaxID=38251;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VIRULENT B;  
RX MEDLINE=96010229; PubMed=7571426;  
RA Zadori Z., Stefancsik R., Rauch T., Kisary J.;  
RT "Analysis of the complete nucleotide sequences of goose and muscovy duck parvoviruses indicates common ancestral origin with adeno-

RT associated virus 2.";  
RL Virology 212:562-573 (1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VIRULENT B;  
RA Zadori Z.;  
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U25749; AAA83230.1; -.  
DR InterPro; IPR001403; Parvo\_coat.  
DR Pfam; PF00740; Parvo\_coat; 1.  
SQ SEQUENCE 732 AA; 81340 MW; 339507C61D47B52C CRC64;  
  
Query Match 52.8%; Score 1717; DB 12; Length 732;  
Best Local Similarity 53.0%; Pred. No. 9.6e-115;  
Matches 327; Conservative 94; Mismatches 148; Indels 48; Gaps 13;  
  
QY 1 TAPGKR-----PVEQSPQ-EPDSSSGIGTGGQPAKRLNFGQDSESVPDPQL 51  
DB 146 TAPAKNTGKLTDFHYPVKKPKLTBEVSAGGSSAVQDG-----GATAE----- 189  
  
QY 52 GEPATPAVGPPTMASGGAPMADNNEGADGVNAGNMHCDSTWLGDVITTTSTRTWA 111  
DB 190 GTEP-----VAASEMAEGGGAMGDSGSGADGVNAGNMHCDSTWMTVTTKTRTW 244  
  
QY 112 LPTNNHLKYQISSASTGASNDNH--YFGYSTPWGFEDENRFHCHFSPRDQRLINNMG 169  
DB 245 LPSYNNHIYKAITS---GTSQDANVQAGYSTPWGYPDENRFHCHFSPRDQRLINNMG 301  
  
QY 170 FRPKRLNFKLFNIQKEVTNDGVTIANNLTSIVQFSDSEYQLPYVLSAHQGLPPF 229  
DB 302 IRPKSLKFKIFNVQKEVTTQDQTKTIANNLSTIIQVFTDDEHQLPYVLSATEGTMPPF 361  
  
QY 230 PADVFMIPQGYLTIN---NGSQAVGRSSFYCLEYFSPQMLRTGNFTSYTBEVPHS 286  
DB 362 PSDVYALPQYGYCTMHTNNGARFNDRSAFYCLEYFSPQMLRTGNFTSYTBEVPHS 421  
  
QY 287 SYAHSQSLDRMLNPLIDQYLYLNRTQNGSGAQNKDLFSRGSFAGMSVQPKNWLPGPC 346  
DB 422 MEAHSQDLRLMNPLOYQYLMNFNEV-DSSRNAQ-----FKKAVKAGYTMGRNWLPGPK 475  
  
QY 347 YRQQRV-SKTCTDNNSNFT-WTGASKYNLNGRESIINPGTAMASHKDEDEKFFMSGVM 404  
DB 476 FLQQRVRAVYGTGTNYANWNINWSNGKNVCLKDQYLLQPGVSATYTEGEASSLPAQNIL 535  
  
QY 405 IFGKE--SAGASNTALDNVMTDEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAM 462  
DB 536 GIADKPYRSGSTAGISDIMWTEBQVAPTINGVWKPYGRTVTNEQNTTAPTSDDLVL 595  
  
QY 463 GALPGMWQDRVYLOGPIWAKIPHTDGHFHPSPMLMGFGLKNPPQILIKNTVPANPP 522  
DB 596 GALPGMWQNDIYLOGPIGAKIPKTDGKFHPSPNLGGFGLHNPPOVFIKNTVPADPP 655  
  
QY 523 AEFSAATKFAFITQYSTGQVSVEIEMELQKENSKRWNPEVOYTSNYAKSANDFTVDNMG 582  
DB 656 VEYVHQKNSYITQYSTGQCTVEVMWELRKENSKRWNPEIQFTSNFSNRTSIMFAPNETG 715  
  
QY 583 LYTEPRPIGTRYLTRPL 599  
DB 716 GYVEDRLIGTRYLTQNL 732  
  
RESULT 10  
Q83290 PRELIMINARY; PRT; 732 AA.  
ID Q83290;  
AC Q83290;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)  
DE Capsid protein.  
OS Muscovy duck parvovirus.  
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.  
OX NCBI\_TaxID=37325;  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN=89384/France;  
RX MEDLINE=96406928; PubMed=8811015;  
RA ie Gall-Recul G., Jestin V., Chagnaud P., Blanchard P., Jestin A.;  
RT Expression of muscovy duck parvovirus capsid proteins (VP2 and VP3)  
RT in a baculovirus expression system and demonstration of immunity  
RT induced by the recombinant protein.";  
RL J. Gen. Virol. 77:2159-2163 (1996).  
DR EMBL; Z68272; CAA92575.1; -.  
DR InterPro; IPR001403; Parvo\_coat.  
DR Pfam; PF00740; Parvo\_coat; 1.  
FT CHAIN 146 732 VP2 CAPSID PROTEIN.  
FT 199 732 VP3 CAPSID PROTEIN.  
SQ SEQUENCE 732 AA; 81364 MW; DE70DCCAB215F4E2 CRC64;  
  
Query Match 52.7%; Score 1714; DB 12; Length 732;  
Best Local Similarity 53.3%; Pred. No. 1.6e-114;  
Matches 327; Conservative 82; Mismatches 164; Indels 40; Gaps 11;  
  
QY 6 KRPEVGEQPEPDSSSGIGK-TGQOPAKKRLNFGQDSESVPDPQLG-----E 53  
DB 141 EEPVNTAPAKSS---GKLTDFHYPIVKKPKLSE---ENSPSPNSGGEASAAATEGSE 192  
  
QY 54 PPATPAVGPPTMASGGAPMADNNEGADGVNAGNMHCDSTWLGDVITTTSTRTWALP 113  
DB 193 PVAAP-----NMAEGSGAGMGDSAGADGVNAGNMHCDSTWLDVTITKTRTWLP 246  
  
QY 114 TYNNHLKYQISSASTGASNDNHFGYSTPWGYPDENRFHCHFSPRDQRLINNMGFRPK 173  
DB 247 SYNHHIYKAITSCTNPDSN-TQYAGYSTPWGYPDENRFHCHFSPRDQRLINNHWGIRPK 305  
  
QY 174 RLNFKLFNIQKEVTNDGVTIANNLTSIVQFSDSEYQLPYVLSAHQGLPPFADV 233  
DB 306 ALKFKIFNVQKEVTTQDQTKTIANNLTSIIQFTDNEHQLPYVLSATEGTMPPPSDV 365  
  
QY 234 FMIPQGYLTIN---NGSQAVGRSSFYCLEYFSPQMLRTGNFTSYTBEVPHSVAH 290  
DB 366 YALPQYGYCTMHTNNGARFNDRSAFYCLEYFSPQMLRTGNFTSYTBEVPHSMFAH 425  
  
QY 291 SQSLDRMLNPLIDQYLYLNRTQNGSGAQNKDLFSRGSFAGMSVQPKNWLPGPCYRQ 350  
DB 426 SQDLRLMNPLOYQYLMNFSEV-NGRNAQ-----FKKAVKAGFAGMGRNWLPGPKLDD 479  
  
QY 351 RVSKTK--TDNNSNFTWTGASKYNLNGRESIINPGTAMASHKDEDEKFFMSGVMIFGK 408  
DB 480 RVRAVSGGTNYANWSIWSKGNVFLKDRYLLQPGVATHTTEDQASSVPAQNITGIK 539  
  
QY 409 E--SAGASNTALDNVMTDEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMALP 466  
DB 540 DPYRSGSTAGISDIMWTEQETIAPTINGVWRPYGLTVTNEQNTTAPTAELLEVGLALP 599  
  
QY 467 GMWQDRVYLOGPIWAKIPHTDGHFHPSPMLMGFGLKNPPQILIKNTVPANPPAEFS 526  
DB 600 GMWQNDIYLOGPIWAKIPKTDGKFHPSPNLGGFGLHNPPOVFIKNTVPADPPLEYV 659  
  
QY 527 ATKFAFITQYSTGQVSVEIEMELQKENSKRWNPEVOYTSNYAKSANDFTVDNNGLYTE 586  
DB 660 NQKNSYITQYSTGQCTVEVMWELRKENSKRWNPEIQFTSNFGNRTSTMFAPNETGYVE 719  
  
QY 587 PRPIGTRYLTRPL 599  
DB 720 DRLIGTRYLTQNL 732  
  
RESULT 11  
Q67667 PRELIMINARY; PRT; 587 AA.  
ID Q67667;  
AC Q67667;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)  
DE VP2.  
GN VP2.

OS goose parvovirus.  
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.  
OX NCBI\_TaxID=38251;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VIRULENT B;  
RX MEDLINE=96010229; PubMed=7571426;  
RA Zadori Z., Stefancsik R., Rauch T., Kisary J.;  
RT "Analysis of the complete nucleotide sequences of goose and muscovy  
RT duck parvoviruses indicates common ancestral origin with adeno-  
RT associated virus 2."  
RL Virology 212:562-573(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VIRULENT B;  
RA Zadori Z.;  
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U25749; AAA83231.1; -.  
DR InterPro; IPR001403; Parvo\_coat.  
DR Pfam; PF00740; Parvo\_coat; 1.  
SQ SEQUENCE 587 AA; 65246 MW; 027BE5D5FE7F0423 CRC64;

Query Match 52.7%; Score 1712; DB 12; Length 587;  
Best Local Similarity 52.9%; Pred. No. 1.6e-114;  
Matches 326; Conservative 94; Mismatches 148; Indels 48; Gaps 13;

OY 2 APGKKR-----PVEQSPQ-EPDSSSGIGKTGQOPAKKRLNFGQTGDSSEVPDPQLG 52  
DB 2 APAKKNKTGKLTLDHYPVKKPKLTEEVSAGGSSAVQD-----GATAE-----G 45  
OY 53 EPPATPAVGPPTMASGGGAPMADNNEGADVGNASGNWHCDSTWLGDRVITSTRTWA 112  
DB 46 TEP-----VAASEMAEGGGAGMGDSGGGADGVGNASGNWHCDSQMGNTVITKTRTW 100  
OY 113 PTYNHLYKQISSASTGASNDN--YFGYSTPWGYPDFNRFCHFSPRDWQRLINNMWG 170  
DB 101 PSYNNHIYKAITS--GTSQDANVOYAGYSTPWGYPDFNRFCHFSPRDWQRLINNMWG 157  
OY 171 RPKRLNFKLNIQVKEVTNDGVTIANNLSTVQVFSDEYQLPYVLSAHQGLPPF 230  
DB 158 RPKSLKFKIFNVQVKEVTTQDKTIANNLTSTIQVFTDEHQLPYVLSAATEGTWPFP 217  
OY 231 ADVFMIPOYGYLTIN--NGSQAVGRSSFYCLEYFPQMLRTGNNFTFSYTEEVPFHS 287  
DB 218 SDVYALPOYGYCTMHTNONGARFNDRSAFYCLEYFPQMLRTGNNFEFTDFEEVPFHS 277  
OY 288 YAHQSQSLDRMLNPLIDQYLYLNRTQNGSGAQNKDLFSRGSAPGMSVQPKNWLPGPC 347  
DB 278 FAHSQDLDRMLNPLVDQYLMNFNEV-DSSRNAQ-----FKKAVKAYGTWGRNWLPGPK 331  
OY 348 RQQRV-SKTKTDNNNSNFT-WTGASKYNLNGRESINPGTAMASHKDEDEKFFPMGVM 405  
DB 332 LDQRVRAVYTGTDNYANWNINWSNGKNVNLKDRQYLLQPGPVSATYTEGEASSLPAQNIL 391  
OY 406 FGKE--SAGASNTALDNVMTDEEIKATNPVATERFGTVAVNFQSSSTDPATGVDHAM 463  
DB 392 IAKDPYRSGSTTAGISDIMVTEGEVAPNGVGWKPYGRTVTNEQNTTAPTSSDLVGL 451  
OY 464 ALPGMWQDRDVLVLOGPIWAKIPHTDGHFSPPLMGFGKLPKPPQILIKNTVPANPPA 523  
DB 452 ALPGMWQNRDIYLOGPIGAKIPKTDGKHPSPLNGFGFLHNPPOVFIKNTVPADPPV 511  
OY 524 EFSATKFAFITQYSTGVSEIEMELQKNSKRWNPEVQYTSNYAKSANVDFVDNNG 583  
DB 512 EYVHQKNSYITQYSTGQCTVEVMWELRKNSKRWNPEIQFTSNFSDRTSIMFAPNETG 571  
OY 584 YTEPRPIGTRYLTRPL 599  
DB 572 YVEDRLIGTRYLTQNL 587

RESULT 12  
Q8V395

ID Q8V395 PRELIMINARY; PRT; 732 AA.  
AC Q8V395;  
DT 01-MAR-2002 (Tremblrel. 20, Created)  
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)  
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
DE Capsid protein VP.  
GN VP.  
OS goose parvovirus.  
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.  
OX NCBI\_TaxID=38251;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=GpV-YG;  
RA Ge Y., You Y., Xu Q.;  
RT "Analysis of the major open reading frames' nucleotide sequences in  
RT Goose parvovirus GpV-YG strain isolated in China."  
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF416726; AAL37722.1; -.  
DR InterPro; IPR001403; Parvo\_coat.  
DR Pfam; PF00740; Parvo\_coat; 1.  
SQ SEQUENCE 732 AA; 81456 MW; 73FE24BC769744B6 CRC64;

Query Match 52.7%; Score 1712; DB 12; Length 732;  
Best Local Similarity 53.2%; Pred. No. 2.2e-114;  
Matches 328; Conservative 89; Mismatches 152; Indels 48; Gaps 13;

OY 1 TAPGKKR-----PVEQSPQ-EPDSSSGIGKTGQOPAKKRLNFGQTGDSSEVPDPQL 51  
DB 146 TAPYKKNKTGKLTLDHYPVKKPKLTEEVSAGGTSVQD-----GATAE----- 189  
OY 52 EPPATPAVGPPTMASGGGAPMADNNEGADVGNASGNWHCDSTWLGDRVITSTRTWA 111  
DB 190 GTEP-----VAASEMAEGGGAGMGDSGGGADGVGNASGNWHCDSQMGNTVITKTRTW 244  
OY 112 LPTYNHLYKQISSASTGASNDN--HYFGYSTPWGYPDFNRFCHFSPRDWQRLINNMWG 169  
DB 245 LPSYNNHIYKAITS--GTSQDATVOYAGYSTPWGYPDFNRFCHFSPRDWQRLINNMWG 301  
OY 170 RPKRLNFKLNIQVKEVTNDGVTIANNLSTVQVFSDEYQLPYVLSAHQGLPPF 229  
DB 302 IRPKSLKFKIFNVQVKEVTTQDKTIANNLTSTIQVFTDEHQLPYVLSAATEGTWPFP 361  
OY 230 PADVMIPOYGYLTIN--NGSQAVGRSSFYCLEYFPQMLRTGNNFTFSYTEEVPFHS 286  
DB 362 PSDVYALPOYGYCTMHTNONGARFNDRSAFYCLEYFPQMLRTGNNFEFTDFEEVPFHS 421  
OY 287 SYAHQSQSLDRMLNPLIDQYLYLNRTQNGSGAQNKDLFSRGSAPGMSVQPKNWLPGPC 346  
DB 422 MFHSQDLDRMLNPLVDQYLMNFNEV-DSSRKAQ-----FKKAVKAYGTWGRNWLPGPK 475  
OY 347 YRQQRV-SKTKTDNNNSNFT-WTGASKYNLNGRESINPGTAMASHKDEDEKFFPMGVM 404  
DB 476 LDQRVRAVYTGTDNYANWNINWSNGKNVNLKDRQYLLQPGPVSATHTTKEASSIPQNIL 535  
OY 405 IFGKE--SAGASNTALDNVMTDEEIKATNPVATERFGTVAVNFQSSSTDPATGVDHAM 462  
DB 536 GLAKDPYRSGSTTAGISDIMVTEGEVAPNGVGWKPYGRTVTNEQNTTAPTSSDLVGL 595  
OY 463 GALPGMWQDRDVLVLOGPIWAKIPHTDGHFSPPLMGFGKLPKPPQILIKNTVPANPPA 522  
DB 596 GALPGMWQNRDIYLOGPIWAKIPKTDGKHPSPLNGFGFLHNPPOVFIKNTVPADPPV 655  
OY 523 AFSATKFAFITQYSTGVSEIEMELQKNSKRWNPEVQYTSNYAKSANVDFVDNNG 582  
DB 656 VEYVHQKNSYITQYSTGQCTVEVMWELRKNSKRWNPEIQFTSNFSDRTSIMFAPNETG 715  
OY 583 YTEPRPIGTRYLTRPL 599  
DB 716 GYIEDRLIGTRYLTQNL 732

RESULT 13  
Q65444



ID	Q65444	PRELIMINARY;	PRT;	732 AA.
AC	Q65444;			
DT	01-NOV-1996 (Tremblrel. 01, Created)			
DT	01-NOV-1996 (Tremblrel. 01, Last sequence update)			
DT	01-JUN-2001 (Tremblrel. 17, Last annotation update)			
DE	Capsid protein VP.			
GN	VP.			
OS	Barbarie duck parvovirus.			
OC	Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.			
OX	NCBI_TaxID=39118;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=FM;			
RX	MEDLINE=96010229; PubMed=7571426;			
RA	Zadori Z., Stefancsik R., Rauch T., Kisary J.;			
RT	"Analysis of the complete nucleotide sequences of goose and muscovy			
RT	duck parvoviruses indicates common ancestral origin with adeno-			
RT	associated virus 2."			
RL	Virology 212:562-573(1995).			
DR	EMBL; U22967; AAA83225.1; -.			
DR	InterPro; IPR001403; Parvo_coat.			
DR	Pfam; PF00740; Parvo_coat; 1.			
DR	SEQUENCE 732 AA; 81314 MW; D92360596E2D2C05 CRC64;			
SO				
Query Match	52.4%; Score 1702; DB 12; Length 732;			
Best Local Similarity	53.0%; Pred. No. 1.1e-113;			
Matches	325; Conservative 83; Mismatches 165; Indels 40; Gaps 11;			
OY	6 KRPEQSPQEPDSSSGIGK-TGQOPAKRLNFGQTGDSSEVPDQPLG-----E 53			
DB	141 EEPVNTAPAKKS-----GKLTDHDPVKKPKLSE---ENSPSPNSNGGEASAAATEGSE 192			
OY	54 PPATPAVGPPTTMAAGGAPMADNNEGADGVGNASGNWCHDSTWLGDRVITTTSTRTWALP 113			
DB	193 PVAAP-----NMAEGSGAMGDSAGADGVGNASGNWCHDSTWLGDRVITTTSTRTWALP 246			
OY	114 TYNNHLYKQISSASTGASNDNHFGYSTPWGYFDNFRFHCHFSPRDWQRLINNHWGFRPK 173			
DB	247 SYNHHYQAITSGTNPDSN-TQYAGYSTPWGYFDNFRFHCHFSPRDWQRLINNHWGIRPK 305			
OY	174 RLNFKLNIQVEKVTNDGVTTIANNLTSTVQVFSDESEYQLPVYLSAHQGLPPEPADV 233			
DB	306 ALKFKEFNQVKEVTQDQTKTIANNLTSTIQIFTDNEHQLPYVLSATEGTMPPEPSDV 365			
OY	234 FMIPQGYLTLN--NGSQAVGRSSFYCLEYFPQSOMLRGNNFTFSYTFEEVPPHSSYAH 290			
DB	366 YALPQGYCTMHTNOSGARFNDRSAFYCLEYFPQSOMLRGNNFEFSFEFEVPPHSMFAH 425			
OY	291 SOSLDRLMNPILDQYLYLNRTONOSGSAQNKDLFSRGSPPAGMSVQPKNMLPGPCYRQO 350			
DB	426 SQDLRLMNPILDQYLMNFSEV-NGRNAQ-----FKKAVKGAFGAMGRNMLPGPKLIDQ 479			
OY	351 RVSKTK--TDNNSNFTWTGASKYNLNGRESIINPGTAMASEKDEDEKFFPMGVMIFGK 408			
DB	480 RVRAYSGGTDNVANWSIWSKGNKVFLLKDRREYLLQPGPVATTHTEDQASSVPAQNITIGIAK 539			
OY	409 E--SAGASNTALDNVMTDEEEKATNPVATERFGTVAVNFQSSSTDPAATGDVHAMGALP 466			
DB	540 DPYRSGSTLAGISDIMVTDEQEIAPTNGVGWRPYGLTVTNEQNTTAPTNAELEVLGALP 599			
OY	467 GMVWQDRDVIYLOGPIWAKIPHTDGHFHPSPMLMGFGGLKNPPQILLIKNTPVPANPPAEFS 526			
DB	600 GMVWQNRDIYLOGPIWAKIPKTDGKPHSPNLGGFGLHNPPOVFIKNTPVPADPLEYV 659			
OY	527 ATKFASFITQYSTGQVSEIEIEMELQKENSKRWNPEVQYTSNYAKSANDFTVDNNGLYTE 586			
DB	660 NQKWNSTYTQYSTGQCTVEVMWELRKENSKRWNPEIQFTSNFGNRTSTMFAPNETGYVE 719			
OY	587 PRPIGTRYLTRPL 599			
DB	720 DRLIGTRYLTQNL 732			

RESULT 14	Q83289	PRELIMINARY;	PRT;	732 AA.
ID	Q83289			
AC	Q83289;			
DT	01-NOV-1996 (Tremblrel. 01, Created)			
DT	01-NOV-1996 (Tremblrel. 01, Last sequence update)			
DT	01-DEC-2001 (Tremblrel. 19, Last annotation update)			
DE	VP1 protein.			
GN	VP1.			
OS	Muscovy duck parvovirus.			
OC	Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.			
OX	NCBI_TaxID=37325;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Zadori Z., Erdei J., Nagy J., Kisari J.;			
RL	Submitted (SEP-1993) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; X75093; CAA52984.1; -.			
DR	InterPro; IPR001403; Parvo_coat.			
DR	Pfam; PF00740; Parvo_coat; 1.			
DR	SEQUENCE 732 AA; 81284 MW; 9FF6955BC15DF3AA CRC64;			
SO				
Query Match	52.3%; Score 1701; DB 12; Length 732;			
Best Local Similarity	53.0%; Pred. No. 1.4e-113;			
Matches	325; Conservative 83; Mismatches 165; Indels 40; Gaps 11;			
OY	6 KRPEQSPQEPDSSSGIGK-TGQOPAKRLNFGQTGDSSEVPDQPLG-----E 53			
DB	141 EEPVNTAPAKKS-----GKLTDHDPVKKPKLSE---ENSPSPNSNGGEASAAATEGSE 192			
OY	54 PPATPAVGPPTTMAAGGAPMADNNEGADGVGNASGNWCHDSTWLGDRVITTTSTRTWALP 113			
DB	193 PVAAP-----NMAEGSGAMGDSAGADGVGNASGNWCHDSTWLGDRVITTTSTRTWALP 246			
OY	114 TYNNHLYKQISSASTGASNDNHFGYSTPWGYFDNFRFHCHFSPRDWQRLINNHWGFRPK 173			
DB	247 SYNHHYQAITSGTNPDSN-TQYAGYSTPWGYFDNFRFHCHFSPRDWQRLINNHWGIRPK 305			
OY	174 RLNFKLNIQVEKVTNDGVTTIANNLTSTVQVFSDESEYQLPVYLSAHQGLPPEPADV 233			
DB	306 ALKFKEFNQVKEVTQDQTKTIANNLTSTIQIFTDNEHQLPYVLSATEGTMPPEPSDV 365			
OY	234 FMIPQGYLTLN--NGSQAVGRSSFYCLEYFPQSOMLRGNNFTFSYTFEEVPPHSSYAH 290			
DB	366 YALPQGYCTMHTNOSGARFNDRSAFYCLEYFPQSOMLRGNNFEFSFEFEVPPHSMFAH 425			
OY	291 SOSLDRLMNPILDQYLYLNRTONOSGSAQNKDLFSRGSPPAGMSVQPKNMLPGPCYRQO 350			
DB	426 SQDLRLMNPILDQYLMNFSEV-NGRNAQ-----FKKAVKGAFGAMGRNMLPGPKLIDQ 479			
OY	351 RVSKTK--TDNNSNFTWTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMGVMIFGK 408			
DB	480 RVRAYSGGTDNVANWSIWSKGNKVFLLKDRREYLLQPGPVATTHTEDQASSVPAQNITIGIAK 539			
OY	409 E--SAGASNTALDNVMTDEEEKATNPVATERFGTVAVNFQSSSTDPAATGDVHAMGALP 466			
DB	540 DPYRSGSTLAGISDIMVTDEQEIAPTNGVGWRPYGLTVTNEQNTTAPTNAELEVLGALP 599			
OY	467 GMVWQDRDVIYLOGPIWAKIPHTDGHFHPSPMLMGFGGLKNPPQILLIKNTPVPANPPAEFS 526			
DB	600 GMVWQNRDIYLOGPIWAKIPKTDGKPHSPNLGGFGLHNPPOVFIKNTPVPADPLEYV 659			
OY	527 ATKFASFITQYSTGQVSEIEIEMELQKENSKRWNPEVQYTSNYAKSANDFTVDNNGLYTE 586			
DB	660 NQKWNSTYTQYSTGQCTVEVMWELRKENSKRWNPEIQFTSNFGNRTSTMFAPNETGYVE 719			
OY	587 PRPIGTRYLTRPL 599			
DB	720 DRLIGTRYLTQNL 732			
RESULT 15	Q9YIJI	PRELIMINARY;	PRT;	724 AA.
ID	Q9YIJI			







GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 2003, 18:20:14 ; Search time 37.1429 Seconds  
(without alignments)  
1915.734 Million cell updates/sec

Title: US-09-807-802A-17

Perfect score: 2906

Sequence: 1 MASGGAPMADNNEGADGVG.....NNGLYTEPRPIGTTRYLTRPL 534

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_101002:\*

1: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1980.DAT:\*  
2: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1981.DAT:\*  
3: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1982.DAT:\*  
4: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1983.DAT:\*  
5: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1984.DAT:\*  
6: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1985.DAT:\*  
7: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1986.DAT:\*  
8: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1987.DAT:\*  
9: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1988.DAT:\*  
10: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1989.DAT:\*  
11: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1990.DAT:\*  
12: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1991.DAT:\*  
13: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1992.DAT:\*  
14: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1993.DAT:\*  
15: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1994.DAT:\*  
16: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1995.DAT:\*  
17: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1996.DAT:\*  
18: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1997.DAT:\*  
19: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1998.DAT:\*  
20: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1999.DAT:\*  
21: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA2000.DAT:\*  
22: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA2001.DAT:\*  
23: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2906	100.0	534	21	AAV71169 Adeno-associated v
2	2906	100.0	599	21	AAV71168 Adeno-associated v
3	2906	100.0	736	21	AAV71167 Adeno-associated v
4	2884	99.2	736	22	AAB59847 AAV6 capsid protei
5	2530	87.1	736	22	AAB59846 AAV3B capsid prote
6	2528	87.0	736	22	AAB59845 AAV3A capsid prote
7	2481.5	85.4	533	22	AA65794 Adeno-associated v
8	2481.5	85.4	598	22	AA65793 Adeno-associated v
9	2481.5	85.4	598	22	AA65793 Adeno-associated v
10	2481.5	85.4	735	22	AA65792 Adeno-associated v

11	2481.5	85.4	735	22	AA651508 Adeno-associated v
12	2481.5	85.4	735	22	AAB59844 AAV2 capsid protei
13	2477.5	85.3	533	22	AA651510 Adeno-associated v
14	2455	84.5	734	22	AAB50326 Adeno-associated v
15	1690.5	58.2	544	19	AAW46314 AAV4 VP3 coat prot
16	1690.5	58.2	598	19	AAW46313 AAV4 VP2 coat prot
17	1690.5	58.2	734	19	AAW46308 AAV4 VP1 capsid pr
18	1690	58.2	736	18	AAW19000 Duck parvovirus ca
19	1676	57.7	534	16	AAR85386 Barbary duck parvo
20	1676	57.7	732	16	AAR85385 Barbary duck parvo
21	1665	57.3	532	21	AAV58162 Adeno associated v
22	1665	57.3	532	23	AAV11407 Adeno associated v
23	1665	57.3	588	23	AAV58161 Adeno associated v
24	1665	57.3	588	23	AAV11406 Adeno associated v
25	1665	57.3	724	21	AAV58160 Adeno associated v
26	1665	57.3	724	23	AAV11405 Adeno-associated v
27	491.5	16.9	756	21	AAV71231 Capsid protein enc
28	490	16.9	554	16	AAW08987 Human parvovirus V
29	490	16.9	554	20	AAV23230 Erythrovirus V9 VP
30	490	16.9	781	16	AAW08986 Human parvovirus V
31	490	16.9	781	20	AAV23227 Erythrovirus V9 VP
32	476.5	16.4	543	12	AAV13405 Parvo virus B19 VP
33	304.5	10.5	264	12	AAV13407 Parvo virus B19 PA
34	295.5	10.2	370	12	AAV13406 Parvo virus B19 PA
35	234.5	8.1	579	13	AAV29079 Porcine Parvovirus
36	228.5	7.9	686	5	AAV40068 Sequence of a porc
37	209.5	7.2	579	14	AAV38697 PSY875 swine parvo
38	205.5	7.1	579	14	AAV38702 Swine parvovirus B
39	205.5	7.1	579	17	AAV39721 Swine parvovirus B
40	205.5	7.1	598	8	AAV70500 Pig parvo virus B
41	200.5	6.9	579	10	AAV94798 Swine parvovirus B
42	198	6.8	584	14	AAV30811 Feline parvovirus
43	194	6.7	584	22	AAE04302 Feline panleukopen
44	189	6.5	584	22	AAV50109 Feline parvovirus
45	188	6.5	584	23	AAV76051 Canine distemper v

ALIGNMENTS

RESULT 1	
AAV71169	AAV71169 standard; Protein; 534 AA.
XX	
AC	AAV71169;
XX	
DT	08-SEP-2000 (first entry)
XX	
DE	Adeno-associated virus serotype 1 capsid protein VP3.
XX	
KW	Adeno-associated virus serotype 1; AAV-1; rep protein; capsid protein;
KW	cap protein; recombinant viral vector; gene delivery; gene therapy;
KW	vaccine; transgene; VP3.
XX	
OS	Adeno associated virus serotype 1.
XX	
PN	WO200028061-A2.
XX	
PD	18-MAY-2000.
XX	
PF	02-NOV-1999; 99WO-US25694.
XX	
PR	05-NOV-1998; 98US-0107114.
XX	
PA	(TYPE-) UNIV PENNSYLVANIA.
XX	
PI	Wilson JM, Xiao W;
XX	
DR	WPI; 2000-376571/32.
DR	N-PSDB; AAD00772, AAD00779.
XX	
PT	Novel adeno-associated virus serotype 1 polynucleotide useful for preparation of medicament for delivery of a transgene to a host

XX Claim 7; Page 99-101; 108pp; English.  
PS  
XX  
CC The patent discloses an adeno-associated virus serotype 1 (AAV-1) DNA  
CC which is characterised by two inverted terminal repeats (ITR) and open  
CC reading frames for rep and capsid (cap) proteins. The rep reading frame  
CC encodes four proteins, Rep 78, Rep 68, Rep 52 and Rep 40, while the cap  
CC reading frame encodes three structural proteins, VP1, VP2 and VP3.  
CC The AAV-1 sequence or its fragments particularly ITRs, rep and cap  
CC coding regions, are useful in production of recombinant viral vectors  
CC for gene delivery. These vectors can be used as gene therapy  
CC vectors, vaccine vectors or antisense delivery vectors. The AAV-1 does  
CC not induce the formation of neutralising antibodies specific to any  
CC serotype of AAV hence is useful for transforming host cells, and in the  
CC preparation of a medicament for the delivery of transgene to a host.  
CC The present sequence is an AAV-1 cap protein VP3 which is  
CC useful in the production of recombinant viral vector for gene delivery.

XX Sequence 534 AA;

Query Match 100.0%; Score 2906; DB 21; Length 534;  
Best Local Similarity 100.0%; Pred. No. 8.4e-237;  
Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASGGAPMADNNEGADVGNASGNWCHDSTWLGDRVITTTSTRTWALPTYNNHLKYQISS 60  
DB 1 MASGGAPMADNNEGADVGNASGNWCHDSTWLGDRVITTTSTRTWALPTYNNHLKYQISS 60  
QY 61 ASTGASNDNHFFGYSTPMGYFDENRFHCHFSPRDWORLNNNWGFRPKRLNFKLFNIQVK 120  
DB 61 ASTGASNDNHFFGYSTPMGYFDENRFHCHFSPRDWORLNNNWGFRPKRLNFKLFNIQVK 120  
QY 121 EVTTNDGVTTIANNLTSTVQVFSDEYQLPYVLSAHQGCLEPPPADVEMIPQGYLTIN 180  
DB 121 EVTTNDGVTTIANNLTSTVQVFSDEYQLPYVLSAHQGCLEPPPADVEMIPQGYLTIN 180  
QY 181 NGSQAVGRSSFFCYCLEYFSPQMLRTGNNFTFSYTFEEVPHSSYAHQSQSLDRLMNPIDQY 240  
DB 181 NGSQAVGRSSFFCYCLEYFSPQMLRTGNNFTFSYTFEEVPHSSYAHQSQSLDRLMNPIDQY 240  
QY 241 LYYLNRTQONQSGSAQNKDLFSRGSFAGMSVQPKNWLPGPCYRQORVSKTKTDNNNSNFT 300  
DB 241 LYYLNRTQONQSGSAQNKDLFSRGSFAGMSVQPKNWLPGPCYRQORVSKTKTDNNNSNFT 300  
QY 301 WTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMGVMIFGKESAGASNTALDNVMITD 360  
DB 301 WTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMGVMIFGKESAGASNTALDNVMITD 360  
QY 361 EEEIKATNPVATERFGTVAVNFQSSSTDPATGVDHAMGALPGMWQDRDYYLQGPWAKI 420  
DB 361 EEEIKATNPVATERFGTVAVNFQSSSTDPATGVDHAMGALPGMWQDRDYYLQGPWAKI 420  
QY 421 PHTDGHFHPSPLMGGFGGLKNPPQILIKNTVPANPPAEFSATKFASTITQYSTQVSV 480  
DB 421 PHTDGHFHPSPLMGGFGGLKNPPQILIKNTVPANPPAEFSATKFASTITQYSTQVSV 480  
QY 481 IEMELQKENSKRWNPEVQYTSNYAKSANVDFTVNNGLYTEPRPIGIRYLRPL 534  
DB 481 IEMELQKENSKRWNPEVQYTSNYAKSANVDFTVNNGLYTEPRPIGIRYLRPL 534

RESULT 2  
AA71168  
ID AA71168 standard; Protein; 599 AA.

XX AC AA71168;  
XX DT 08-SEP-2000 (first entry)

DE Adeno-associated virus serotype 1 capsid protein VP2.  
XX Adeno-associated virus serotype 1; AAV-1; rep protein; capsid protein;  
KW cap protein; recombinant viral vector; gene delivery; gene therapy;

KW vaccine; transgene; VP2.  
XX  
OS Adeno associated virus serotype 1.

XX MO200028061-A2.

XX 18-MAY-2000.

XX 02-NOV-1999; 99WO-US25694.

XX 05-NOV-1998; 98US-0107114.

XX (UNIV PENNSYLVANIA.

PI Wilson JM, Xiao W;

DR WPI; 2000-376571/32.

DR N-PSDB; AAD00772, AAD00778.

PT Novel adeno-associated virus serotype 1 polynucleotide useful for  
PT preparation of medicament for delivery of a transgene to a host  
XX  
XX  
PS Claim 7; Page 93-95; 108pp; English.

XX  
CC The patent discloses an adeno-associated virus serotype 1 (AAV-1) DNA  
CC which is characterised by two inverted terminal repeats (ITR) and open  
CC reading frames for rep and capsid (cap) proteins. The rep reading frame  
CC encodes four proteins, Rep 78, Rep 68, Rep 52 and Rep 40, while the cap  
CC reading frame encodes three structural proteins, VP1, VP2 and VP3.  
CC The AAV-1 sequence or its fragments particularly ITRs, rep and cap  
CC coding regions, are useful in production of recombinant viral vectors  
CC for gene delivery. These vectors can be used as gene therapy  
CC vectors, vaccine vectors or antisense delivery vectors. The AAV-1 does  
CC not induce the formation of neutralising antibodies specific to any  
CC serotype of AAV hence is useful for transforming host cells, and in the  
CC preparation of a medicament for the delivery of transgene to a host.  
CC The present sequence is an AAV-1 cap protein VP2 which is  
CC useful in the production of recombinant viral vector for gene delivery.

XX Sequence 599 AA;

Query Match 100.0%; Score 2906; DB 21; Length 599;  
Best Local Similarity 100.0%; Pred. No. 9.9e-237;  
Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASGGAPMADNNEGADVGNASGNWCHDSTWLGDRVITTTSTRTWALPTYNNHLKYQISS 60  
DB 66 MASGGAPMADNNEGADVGNASGNWCHDSTWLGDRVITTTSTRTWALPTYNNHLKYQISS 125  
QY 61 ASTGASNDNHFFGYSTPMGYFDENRFHCHFSPRDWORLNNNWGFRPKRLNFKLFNIQVK 120  
DB 126 ASTGASNDNHFFGYSTPMGYFDENRFHCHFSPRDWORLNNNWGFRPKRLNFKLFNIQVK 185  
QY 121 EVTTNDGVTTIANNLTSTVQVFSDEYQLPYVLSAHQGCLEPPPADVEMIPQGYLTIN 180  
DB 186 EVTTNDGVTTIANNLTSTVQVFSDEYQLPYVLSAHQGCLEPPPADVEMIPQGYLTIN 245  
QY 181 NGSQAVGRSSFFCYCLEYFSPQMLRTGNNFTFSYTFEEVPHSSYAHQSQSLDRLMNPIDQY 240  
DB 246 NGSQAVGRSSFFCYCLEYFSPQMLRTGNNFTFSYTFEEVPHSSYAHQSQSLDRLMNPIDQY 305  
QY 241 LYYLNRTQONQSGSAQNKDLFSRGSFAGMSVQPKNWLPGPCYRQORVSKTKTDNNNSNFT 300  
DB 306 LYYLNRTQONQSGSAQNKDLFSRGSFAGMSVQPKNWLPGPCYRQORVSKTKTDNNNSNFT 365  
QY 301 WTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMGVMIFGKESAGASNTALDNVMITD 360  
DB 366 WTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMGVMIFGKESAGASNTALDNVMITD 425  
QY 361 EEEIKATNPVATERFGTVAVNFQSSSTDPATGVDHAMGALPGMWQDRDYYLQGPWAKI 420  
DB 426 EEEIKATNPVATERFGTVAVNFQSSSTDPATGVDHAMGALPGMWQDRDYYLQGPWAKI 485

QY	421	PHTDGHFHPSPLMGGFGLKNPPQILLIKNTVPANPPAEFSATKFAFITQYSTGVSV	480
DB	486	PHTDGHFHPSPLMGGFGLKNPPQILLIKNTVPANPPAEFSATKFAFITQYSTGVSV	545
QY	481	IEWELQENSKRWNPVEQYTSNYSKASANDFTVDNNGLYTEPRDIGTRYLTRPL	534
DB	546	IEWELQENSKRWNPVEQYTSNYSKASANDFTVDNNGLYTEPRDIGTRYLTRPL	599
RESULT 3			
ID	AAAY71167	standard; Protein; 736 AA.	
XX			
AC	AAAY71167;		
XX			
DT	08-SEP-2000	(first entry)	
XX			
DE	Adeno-associated virus serotype 1 capsid protein VP1.		
XX			
KM	Adeno-associated virus serotype 1; AAV-1; rep protein; capsid protein;		
KW	cap protein; recombinant viral vector; gene delivery; gene therapy;		
KM	vaccine; transgene; VP1.		
XX			
OS	Adeno associated virus serotype 1.		
XX			
PN	WO200028061-A2.		
PD	18-MAY-2000.		
XX			
PF	02-NOV-1999;	99WO-US25694.	
XX			
PR	05-NOV-1998;	98US-0107114.	
XX			
PA	(UYPE-) UNIV PENNSYLVANIA.		
XX			
PI	Wilson JM, Xiao W;		
XX			
DR	WPI: 2000-376571/32.		
XX			
PT	N-PSDB; AAD00772, AAD00777.		
XX			
PT	Novel adeno-associated virus serotype 1 polynucleotide useful for		
PT	preparation of medicament for delivery of a transgene to a host		
XX			
PS	Claim 7; Page 87-90; 108bp; English.		
XX			
CC	The patent discloses an adeno-associated virus serotype 1 (AAV-1) DNA		
CC	which is characterised by two inverted terminal repeats (ITR) and open		
CC	reading frames for rep and capsid (cap) proteins. The rep reading frame		
CC	encodes four proteins, Rep 78, Rep 68, Rep 52 and Rep 40, while the cap		
CC	reading frame encodes three structural proteins, VP1, VP2 and VP3.		
CC	The AAV-1 sequence or its fragments particularly ITRs, rep and cap		
CC	coding regions, are useful in production of recombinant viral vectors		
CC	for gene delivery. These vectors can be used as gene therapy		
CC	vectors, vaccine vectors or antisense delivery vectors. The AAV-1 does		
CC	not induce the formation of neutralising antibodies specific to any		
CC	serotype of AAV hence is useful for transforming host cells, and in the		
CC	preparation of a medicament for the delivery of transgene to a host.		
CC	The present sequence is an AAV-1 cap protein VP1 which is		
CC	useful in the production of recombinant viral vector for gene delivery.		
XX			
XX			
SQ	Sequence	736 AA;	
QY	Query Match	100.0%; Score 2906; DB 21; Length 736;	
DB	Best Local Similarity	100.0%; Pred. No. 1.3e-236;	
DB	Matches 534; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	
QY	1	MASGGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVITTTSTRTWALPTYNNHLYKQISS	60
DB	203	MASGGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVITTTSTRTWALPTYNNHLYKQISS	262
QY	61	ASTGASNDNHYFGYSTPWGYPFDNRFHCHFSPRDWQRLINNNWGRPKRLNFKLFNIQVK	120
DB	263	ASTGASNDNHYFGYSTPWGYPFDNRFHCHFSPRDWQRLINNNWGRPKRLNFKLFNIQVK	322

QY	121	EVTITNDGVTITIANNLSTSTVQVSDSEYQLPYVLGSAHQGCLPPFPADVMIPOYGYLTLN	180
Db	323	EVTITNDGVTITIANNLSTSTVQVSDSEYQLPYVLGSAHQGCLPPFPADVMIPOYGYLTLN	382
QY	181	NGSQAVGRSSFYCLEYFPSSQMLRTGNNFTFSYTFEEVPFHSSYAHQSGLDRIMNPLIDQY	240
Db	383	NGSQAVGRSSFYCLEYFPSSQMLRTGNNFTFSYTFEEVPFHSSYAHQSGLDRIMNPLIDQY	442
QY	241	LYYLNRTQNSGSAQNKDLLFSGSPAGMSVQPKNMLPGPCYRQQRVSKTKTDNNSNFT	300
Db	443	LYYLNRTQNSGSAQNKDLLFSGSPAGMSVQPKNMLPGPCYRQQRVSKTKTDNNSNFT	502
QY	301	WTGASKYNLNGRESIINPGTAMASHKDEDEKFEFPMGVMIFGKESAGASNTALDNVMTD	360
Db	503	WTGASKYNLNGRESIINPGTAMASHKDEDEKFEFPMGVMIFGKESAGASNTALDNVMTD	562
QY	361	EEIKATNPVATERFEGTAAVNFQSSSTD PATGDVHAMGALPGMWODRDVYLQGPWAKI	420
Db	563	EEIKATNPVATERFEGTAAVNFQSSSTD PATGDVHAMGALPGMWODRDVYLQGPWAKI	622
QY	421	PHTDGHFHPSPLMGGFGLKNPPQILIKNTPVPANPPAEFSATKFASTTOYSTGOVSVE	480
Db	623	PHTDGHFHPSPLMGGFGLKNPPQILIKNTPVPANPPAEFSATKFASTTOYSTGOVSVE	682
QY	481	IEMELQKENSKRWNPEVOYTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL	534
Db	683	IEMELQKENSKRWNPEVOYTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL	736
RESULT 4			
ID	AAB59847	standard; Protein; 736 AA.	
XX	AAB59847;		
XX	28-MAR-2001	(first entry)	
XX	AAV6 capsid protein VP1.		
XX	AAV6; gene therapy; AAV viral vector; cystic fibrosis; cancer; AIDS;		
KW	atherosclerosis; sickle cell anaemia; thalassaemia;		
KW	blood clotting disorder; diabetes; capsid protein VP1.		
OS	Adeno associated virus.		
XX	US6156303-A.		
XX	05-DEC-2000.		
XX	11-JUN-1997;	97US-0873168.	
XX	11-JUN-1997;	97US-0873168.	
XX	(UNIW ) UNIV WASHINGTON.		
XX	Russell DW, Rutledge EA;		
XX	WPI; 2001-060164/07.		
PT	Adeno-associated virus serotype 6 and viral vector derived from it for		
PT	gene therapy of cystic fibrosis, cancer, acquired immunodeficiency		
PT	syndrome, sickle cell anemia, thalassemia and diabetes		
XX	Disclosure; Fig 2; 50pp; English.		
XX	The present invention relates to adeno-associated virus serotypes. The		
CC	present sequence is capsid protein VP1 of one such serotype (AAV6). AAV6		
CC	can be used to construct AAV viral vectors for use in gene therapy for a		
CC	range of disorders: cystic fibrosis, cancer, AIDS, atherosclerosis,		
CC	sickle cell anaemia, thalassaemia, blood clotting disorders and diabetes.		
CC	The AAV viral vectors have increased transduction efficiency of a		
CC	particular host cell as the AAV virion containing the AAV vector genome		





KW atherosclerosis; sickle cell anaemia; thalassaemia;  
KW blood clotting disorder; diabetes; capsid protein VP1.  
OS Adeno associated virus.

PN US6156303-A.

PD 05-DEC-2000.

PF 11-JUN-1997; 97US-0873168.

PR 11-JUN-1997; 97US-0873168.

PA (UNIW ) UNIV WASHINGTON.

PI Russell DW, Rutledge EA;

DR WPI; 2001-060164/07.

PT Adeno-associated virus serotype 6 and viral vector derived from it for  
PT gene therapy of cystic fibrosis, cancer, acquired immunodeficiency  
PT syndrome, sickle cell anemia, thalassemia and diabetes

PS Disclosure; Fig 2; 50pp; English.

CC The present invention relates to adeno-associated virus serotypes. The  
CC present sequence is capsid protein VP1 of one such serotype (AAV3A).  
CC AAV3A can be used to construct AAV viral vectors for use in gene therapy  
CC for a range of disorders: cystic fibrosis, cancer, AIDS, atherosclerosis,  
CC sickle cell anaemia, thalassaemia, blood clotting disorders and diabetes.  
CC The AAV viral vectors have increased transduction efficiency of a  
CC particular host cell as the AAV virion containing the AAV vector genome  
CC can be modified to express a capsid protein of an AAV serotype that  
CC transduces the selected host cell.

XX Sequence 736 AA;

Query Match 87.0%; Score 2528; DB 22; Length 736;  
Best Local Similarity 85.8%; Pred. No. 1.2e-204;  
Matches 459; Conservative 28; Mismatches 46; Indels 2; Gaps 2;

OY 1 MASGGGAPMADNNEGADGVGNASGNWHDSTWLGDRVITTSRTWALPTYNNHLYKQISS 60  
DB 203 MASGGGAPMADNNEGADGVGNSSGNWHDSTWLGDRVITTSRTWALPTYNNHLYKQISS 262  
OY 61 ASTGASNDNHFGYSTPWGFDENRHFCHFSPRDWORLNNNGFRPKRLNFKLFNIQVK 120  
DB 263 QS-GASNDNHFGYSTPWGFDENRHFCHFSPRDWORLNNNGFRPKRLNFKLFNIQVK 321  
OY 121 EYTTNDGVTTIANNLTSTVOVFSDEXOLPYVLGSAHQCLPPPADVEMIPQGYLTIN 180  
DB 322 GVTQNDGVTTIANNLTSTVOVFSDEXOLPYVLGSAHQCLPPPADVEMIPQGYLTIN 381  
OY 181 NGSQAVGRSSFYCLEYFSPQMLRTGNFTFSYTFEEVPFHSYAHQSQSLDRMLNPLIDQY 240  
DB 382 NGSQAVGRSSFYCLEYFSPQMLRTGNFTFSYTFEEVPFHSYAHQSQSLDRMLNPLIDQY 441  
OY 241 LYYLNRTO-NQSGSAQNKDLFSRGSAPGMSVQPKWLPGPCYRQORVSKTKTDNNNSNF 299  
DB 442 LYYLNRTOGTTSTTQSRLLFSQAGFQSMQLQARNWLPGPCYRQORLSKTANDNNNSNF 501  
OY 300 TWTGASKYNLNGRESIINPGTAMASHKDEDEKFPFMSGVMIFGESAGASNTALDNVMTT 359  
DB 502 PWTAAASKYHLNGRDSLVPNGPAMASHKDEDEKFPFMSGVMIFGESAGASNTALDNVMTT 561  
OY 360 DEEEIKATNPVATERFGTAVAVNFQSSSTDPATGVDHAMGALPGWVQDRDVLVYLGQPIWAK 419  
DB 562 DEEBIRTNPVATEQYGTAVANNLQSSNTAPTGTGVNHQALPGWVQDRDVLVYLGQPIWAK 621  
OY 420 IPHTDGHFHSPLMGFGGLKPNPQIILKNTVPANPAPAFSATAKFAFITQYSTGVSV 479  
DB 622 IPHTDGHFHSPLMGFGGLKPNPQIILKNTVPANPAPAFSATAKFAFITQYSTGVSV 681

OY 480 EIEWELQENSKRWNPVOYTSNYAKSANDFTVDNNGLYTEPRPIGTRYLTRPL 534  
DB 682 EIEWELQENSKRWNPVOYTSNYAKSANDFTVDNNGLYTEPRPIGTRYLTRPL 736

RESULT 7  
AAG65794

ID AAG65794 standard; Protein; 533 AA.

AC AAG65794;

DT 11-FEB-2002 (first entry)

DE Adeno-associated virus 2 (AAV-2) major coat protein VP3.

KW Recombinant viral vector; RVV; capsid; parvovirus; transgene; cytostatic;  
KW inverted terminal repeat; neurotropic; neuroprotective; antianemic; ITR;  
KW antidiabetic; antitumour; gene therapy; adeno-associated virus; AAV;  
KW major coat protein; AAV-2; VP3.

OS Adeno-associated virus 2.

PN WO200168888-A2.

PD 20-SBP-2001.

PF 13-MAR-2001; 2001WO-US07927.

PR 14-MAR-2000; 2000US-189110P.

PA (NEUR-) NEUROLOGIX INC.

PI Xiao W, During MJ;

DR WPI; 2001-596912/67.

DR N-PSDB; AAI66974.

PT Recombinant viral vector useful in improving gene therapy in a subject,  
PT and for increasing efficiency of entry into a cell, comprises a  
PT chimeric capsid having one non-native amino acid sequence and a desired  
PT transgene

PS Disclosure; Page 51; 53pp; English.

CC The invention provides a recombinant viral vector (RVV) comprising a  
CC chimeric capsid (I) having at least one non-native amino acid sequence,  
CC derived from a capsid protein domain of parvovirus (II), a virus (III),  
CC or their combination, and a transgene flanked 5' and 3' by inverted  
CC terminal repeat (ITR) sequences, derived from (II), (III), or their  
CC combination. The RVV is useful for improving gene therapy in a subject  
CC with a disorder, and for increasing the efficiency of entry into a cell,  
CC which involves producing (I) encapsulating a viral vector, and contacting  
CC a cell with RVV having (I) such that (I) binds to an attachment site on  
CC the cell surface and permits the vector to enter the cell efficiently.  
CC A pharmaceutical composition comprising RVV with (I) containing a  
CC transgene sequence associated with a disease or a disorder such that  
CC expression of the transgene would result in amelioration of the disease  
CC or disorder such as inherited neurological and metabolic diseases e.g.  
CC lysosomal storage disease, Lesch-Nyhan syndrome, amyloid polynuropathy,  
CC Alzheimer amyloid, Duchenne's muscular dystrophy, diseases of the blood  
CC e.g. sickle-cell anemia, clotting disorders and thalassemias, cystic  
CC fibrosis, diabetes, diseases associated with hormone deficiencies,  
CC retinoblastoma and various types of neoplastic cells which include  
CC tumours especially central nervous system tumours, neoplasms, carcinomas,  
CC sarcomas, leukemias and lymphoma. The present sequence represents the  
CC adeno-associated virus 2 (AAV-2) major coat protein VP3. AAV-2 sequences  
CC are used in the construction of a chimeric vector.

SO Sequence 533 AA;

Query Match 85.4%; Score 2481.5; DB 22; Length 533;  
Best Local Similarity 83.3%; Pred. No. 6.2e-201;  
Matches 445; Conservative 37; Mismatches 51; Indels 1; Gaps 1;

QY 1 MASGGAPMADNNEGADGVGNASGNWCHDSTWLGDRVITTTSTRTWALPTYNNHLYKQISS 60  
DB 1 MATGSGAPMADNNEGADGVGNSSGNWCHDSTWMDRVITTTSTRTWALPTYNNHLYKQISS 60  
QY 61 ASTGASNDNHFFGYSTPWGYFDENRFCHFSPRDWQRLINNNWGFRPKRLNFKLFNIQVK 120  
DB 61 QS-GASNDNHFFGYSTPWGYFDENRFCHFSPRDWQRLINNNWGFRPKRLNFKLFNIQVK 119  
QY 121 EVTTNDGVTTIANNLTSTVQVFSDEYQLPYVLGSAHQGLPPFPADVEMIPQGYLTIN 180  
DB 120 EVTQNDGTTTIANNLTSTVQVFTDSEYQLPYVLGSAHQGLPPFPADVEMVPOGYLTIN 179  
QY 181 NSQAVGRSSFFCYLEYFSPQMLRTGNNFTFSYTFEEVPEFHSSYAHQSQSLDRLMPLIDQY 240  
DB 180 NSQAVGRSSFFCYLEYFSPQMLRTGNNFTFSYTFEDVPEFHSSYAHQSQSLDRLMPLIDQY 239  
QY 241 LYYLNRTQNGSQAQNKDLLFSRGSFAGMSVQPKNWLPGPCYRQQRVSKTKTDNNNSFT 300  
DB 240 LYYLSRTNTPSGTTQSRLOFSQAGASDIRDQSRNWLPGPCYRQQRVSKTSADNNSEYS 299  
QY 301 WTGASKYNLNGRESIINPGTAMASHKODEDKFFPMGVMIFGKESAGASNTALDNVMTD 360  
DB 300 WTGATKYHLNGRDSLVPNPGPAMASHKDEKEFFPQSGVLIFGKQSEKTVDIEKVMITD 359  
QY 361 EEEIKATNPVATERFGTVAVNFQSSSTDPATGVHAMGALPGMWQDRDVLVLOGPIMAKI 420  
DB 360 EEEIRTNPVATEQYGSVSTNLQGRNRQAATADVNTQGVLPGMWQDRDVLVLOGPIMAKI 419  
QY 421 PHTDGHFHPSPILMGFGLKNPPQILIKNTPVPANPPAEFSATKFAFITQYSTGQVSVE 480  
DB 420 PHTDGHFHPSPILMGFGLKHPRPQILIKNTPVPANPSTTFSAAKFASFITQYSTGQVSVE 479  
QY 481 IEWELQKENSKRWNPEVQYTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 534  
DB 480 IEWELQKENSKRWNPEIQYTSNYNKS VNVDFTVDNNGVYSEPRPIGTRYLTRNL 533

RESULT 8  
AAG65793  
ID AAG65793 standard; Protein; 598 AA.  
XX AC AAG65793;  
XX DT 11-FEB-2002 (first entry)  
XX DE Adeno-associated virus 2 (AAV-2) major coat protein VP2.  
XX KM Recombinant viral vector; RVV; capsid; parvovirus; transgene; cytosstatic;  
XX KM inverted terminal repeat; neotropic; neuroprotective; antineoplastic; ITR;  
XX KM antidiabetic; antitumour; gene therapy; adeno-associated virus; AAV;  
XX KM major coat protein; AAV-2; VP2.  
OS Adeno-associated virus 2.  
XX FH Key Location/Qualifiers  
FT Misc-difference 1  
FT /note= "encoded by ACG"  
XX PN WO200168888-A2.  
XX PD 20-SEP-2001.  
XX PF 13-MAR-2001; 2001WO-US07927.  
XX PR 14-MAR-2000; 2000US-189110P.  
XX PA (NEUR-) NEUROLOGIX INC.  
XX PI Xiao W, During MJ;  
XX WPI: 2001-596912/67.  
DR N-PSDB; AAI66974.

XX PT Recombinant viral vector useful in improving gene therapy in a subject,  
PT and for increasing efficiency of entry into a cell, comprises a  
PT chimeric capsid having one non-native amino acid sequence and a desired  
PT transgene  
XX PS Disclosure; Page 51; 53pp; English.  
XX The invention provides a recombinant viral vector (RVV) comprising a  
CC chimeric capsid (I) having at least one non-native amino acid sequence,  
CC derived from a capsid protein domain of parvovirus (II), a virus (III),  
CC or their combination, and a transgene flanked 5' and 3' by inverted  
CC terminal repeat (ITR) sequences, derived from (II), (III), or their  
CC combination. The RVV is useful for improving gene therapy in a subject  
CC with a disorder, and for increasing the efficiency of entry into a cell,  
CC which involves producing (I) encapsulating a viral vector, and contacting  
CC the cell with RVV having (I) such that (I) binds to an attachment site on  
CC the cell surface and permits the vector to enter the cell efficiently.  
CC A pharmaceutical composition comprising RVV with (I) containing a  
CC transgene sequence associated with a disease or a disorder such that  
CC expression of the transgene would result in amelioration of the disease  
CC or disorder such as inherited neurological and metabolic diseases e.g.  
CC lysosomal storage disease, Lesch-Nyhan syndrome, amyloid polynuropathy,  
CC Alzheimer amyloid, Duchenne's muscular dystrophy, diseases of the blood  
CC e.g. sickle-cell anemia, clotting disorders and thalassemias, cystic  
CC fibrosis, diabetes, diseases associated with hormone deficiencies,  
CC retinoblastoma and various types of neoplastic cells which include  
CC tumours especially central nervous system tumours, neoplasms, carcinomas,  
CC sarcomas, leukemias and lymphoma. The present sequence represents the  
CC adeno-associated virus 2 (AAV-2) major coat protein VP2. AAV-2 sequences  
CC are used in the construction of a chimeric vector.  
XX SQ Sequence 598 AA;  
QY Query Match 85.4%; Score 2481.5; DB 22; Length 598;  
DB Best Local Similarity 83.3%; Pred. No. 7.3e-201;  
Matches 445; Conservative 37; Mismatches 51; Indels 1; Gaps 1;  
QY 1 MASGGAPMADNNEGADGVGNASGNWCHDSTWLGDRVITTTSTRTWALPTYNNHLYKQISS 60  
DB 66 MATGSGAPMADNNEGADGVGNSSGNWCHDSTWMDRVITTTSTRTWALPTYNNHLYKQISS 125  
QY 61 ASTGASNDNHFFGYSTPWGYFDENRFCHFSPRDWQRLINNNWGFRPKRLNFKLFNIQVK 120  
DB 126 QS-GASNDNHFFGYSTPWGYFDENRFCHFSPRDWQRLINNNWGFRPKRLNFKLFNIQVK 184  
QY 121 EVTTNDGVTTIANNLTSTVQVFSDEYQLPYVLGSAHQGLPPFPADVEMIPQGYLTIN 180  
DB 185 EVTQNDGTTTIANNLTSTVQVFTDSEYQLPYVLGSAHQGLPPFPADVEMVPOGYLTIN 244  
QY 181 NSQAVGRSSFFCYLEYFSPQMLRTGNNFTFSYTFEEVPEFHSSYAHQSQSLDRLMPLIDQY 240  
DB 245 NSQAVGRSSFFCYLEYFSPQMLRTGNNFTFSYTFEDVPEFHSSYAHQSQSLDRLMPLIDQY 304  
QY 241 LYYLNRTQNGSQAQNKDLLFSRGSFAGMSVQPKNWLPGPCYRQQRVSKTKTDNNNSFT 300  
DB 305 LYYLSRTNTPSGTTQSRLOFSQAGASDIRDQSRNWLPGPCYRQQRVSKTSADNNSEYS 364  
QY 301 WTGASKYNLNGRESIINPGTAMASHKODEDKFFPMGVMIFGKESAGASNTALDNVMTD 360  
DB 365 WTGATKYHLNGRDSLVPNPGPAMASHKDEKEFFPQSGVLIFGKQSEKTVDIEKVMITD 424  
QY 361 EEEIKATNPVATERFGTVAVNFQSSSTDPATGVHAMGALPGMWQDRDVLVLOGPIMAKI 420  
DB 425 EEEIRTNPVATEQYGSVSTNLQGRNRQAATADVNTQGVLPGMWQDRDVLVLOGPIMAKI 484  
QY 421 PHTDGHFHPSPILMGFGLKNPPQILIKNTPVPANPPAEFSATKFAFITQYSTGQVSVE 480  
DB 485 PHTDGHFHPSPILMGFGLKHPRPQILIKNTPVPANPSTTFSAAKFASFITQYSTGQVSVE 544  
QY 481 IEWELQKENSKRWNPEVQYTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 534  
DB 545 IEWELQKENSKRWNPEIQYTSNYNKS VNVDFTVDNNGVYSEPRPIGTRYLTRNL 598



RESULT 9  
AAM51509  
ID AAM51509 standard; protein; 598 AA.  
XX  
AC AAM51509;  
XX  
DT 02-JAN-2002 (first entry)  
XX  
DE Adeno-associated virus VP2 capsid protein.  
XX  
KW Adeno-associated virus; AAV; VP2; capsid; virus-like particle;  
KW nuclear localisation signal; VP3.  
XX  
OS Adeno associated virus.  
XX  
PN JP2001169777-A.  
XX  
PD 26-JUN-2001.  
XX  
PF 30-JUL-1999; 99JP-0249140.  
XX  
PR 30-JUL-1999; 99JP-0249140.  
XX  
PA (HAND/) HANDA H.  
XX  
DR WPI; 2001-599854/68.  
XX  
PT New virus-like particles from VP3 capsid protein of adeno-associated  
PT virus, comprise a peptide containing a nucleus-shifting signal  
PT connected to its N-terminal -  
XX  
PS Claim 1; Page 14-16; 33pp; Japanese.  
XX  
CC The present sequence is provided in a specification relating to a  
CC virus-like particle-forming protein, and to a peptide containing a  
CC nuclear-shifting signal at its N-terminus that can form virus-like  
CC particles by shifting to the nucleus of the animal cell in which it  
CC is expressed. The method is used for forming virus-like particles  
CC from the VP3 capsid protein of adeno-associated virus (AAV).  
XX  
SQ Sequence 598 AA;  
  
Query Match 85.4%; Score 2481.5; DB 22; Length 598;  
Best Local Similarity 83.3%; Pred. No. 7.3e-201;  
Matches 445; Conservative 37; Mismatches 51; Indels 1; Gaps 1;  
  
QY 1 MASGGGAPMADNNEGADGVGNASGNWCDSTWLGDRVITSTRTWALPTYNHLYKQISS 60  
DB 66 MATGSGAPMADNNEGADGVGNSSGNWCDSTWMDRVITSTRTWALPTYNHLYKQISS 125  
QY 61 ASTGASNDNHFGYSTPWGFDENRFHCHFSRDMQRLINNNWGFRRPKLNFKLFNIQVK 120  
DB 126 QS-GASNDNHFGYSTPWGFDENRFHCHFSRDMQRLINNNWGFRRPKLNFKLFNIQVK 184  
QY 121 EYTTNDGVTIANNLSTVQVFSDSRYQLPYVLGSAHQGLPPPADVEMIPQYGLTLN 180  
DB 185 EYTONDGTITIANNLSTVQVFTDSEYQLPYVLGSAHQGLPPPADVEMIPQYGLTLN 244  
QY 181 NGSQAVGRSSFYCLEYFSPQMLRTGNFTFSYTFEBVPFHSSYAHQSOLDRMLNPLIDQY 240  
DB 245 NGSQAVGRSSFYCLEYFSPQMLRTGNFTFSYTFEBVPFHSSYAHQSOLDRMLNPLIDQY 304  
QY 241 LYYLNRTONQSGSAQNKDLFSRGSPPAGMSVQPKNWLPGPCYRQORVSKTKTDNNNSNFT 300  
DB 305 LYYLSRTNTPSGTTTQSRLOFSQAGASDIRDQSRNWLPGPCYRQORVSKTSADNNNSEYS 364  
QY 301 WTGASKYVNLNGRESIINPGTAMASHKDEDEKFFPMGVMIFGKESAGASNTALDNVMTD 360  
DB 365 WTGATKYHLNGRDSLVPNGPAMASHKDEDEKFFPQSGVLIFGKQSEKTNVDIEKVMITD 424  
QY 361 EEEIKATNPVATERFGTVAVNFQSSSTDPATGTVHAMGALPGMVQDRDVTYLGPIWAKI 420

DB 425 EEEIRTPNVPATEQYGSVSTNLQGNRQATADAVNTQGVLPGMVWQDRDVTYLGPIWAKI 484  
QY 421 PHTDGHFHPSPMLMGFGFLKNPPQIILIKNTPVPANPPAEFSATKFAFITQYSTGVSE 480  
DB 485 PHTDGHFHPSPMLMGFGFLKNPPQIILIKNTPVPANPSTTSAAKFASFITQYSTGVSE 544  
QY 481 IEWELOKNSKRWNPVEVQYTSNYAKSANVDFTVDDNGLYTEPRPIGTRYLTRPL 534  
DB 545 IEWELOKNSKRWNPQIYTSNYKNSVNVDFTVDTNGVSEPRPIGTRYLTRNL 598  
  
RESULT 10  
AAG65792  
ID AAG65792 standard; Protein; 735 AA.  
XX  
AC AAG65792;  
XX  
DT 11-FEB-2002 (first entry)  
XX  
DE Adeno-associated virus 2 (AAV-2) major coat protein VP1.  
XX  
KW Recombinant viral vector; RVV; capsid; parvovirus; transgene; cytosolic;  
KW inverted terminal repeat; nootropic; neuroprotective; antianemic; ITR;  
KW antidiabetic; antitumour; gene therapy; adeno-associated virus; AAV;  
KW major coat protein; AAV-2; VP1.  
XX  
OS Adeno-associated virus 2.  
XX  
PN WO200168888-A2.  
XX  
PD 20-SEP-2001.  
XX  
PF 13-MAR-2001; 2001WO-US07927.  
XX  
PR 14-MAR-2000; 2000US-189110P.  
XX  
PA (NEUR-) NEUROLOGIX INC.  
XX  
PI Xiao W, During MJ;  
XX  
DR WPI; 2001-596912/67.  
DR N-PSDB; AAI66974.  
XX  
PT Recombinant viral vector useful in improving gene therapy in a subject,  
PT and for increasing efficiency of entry into a cell, comprises a  
PT chimeric capsid having one non-native amino acid sequence and a desired  
PT transgene -  
XX  
PS Disclosure; Page 50; 53pp; English.  
XX  
CC The invention provides a recombinant viral vector (RVV) comprising a  
CC chimeric capsid (I) having at least one non-native amino acid sequence,  
CC derived from a capsid protein domain of parvovirus (II), a virus (III),  
CC or their combination, and a transgene flanked 5' and 3' by inverted  
CC terminal repeat (ITR) sequences, derived from (II), (III), or their  
CC combination. The RVV is useful for improving gene therapy in a subject  
CC with a disorder, and for increasing the efficiency of entry into a cell,  
CC which involves producing (I) encapsulating a viral vector, and contacting  
CC a cell with RVV having (I) such that (I) binds to an attachment site on  
CC the cell surface and permits the vector to enter the cell efficiently.  
CC A pharmaceutical composition comprising RVV with (I) containing a  
CC transgene sequence associated with a disease or a disorder such that  
CC expression of the transgene would result in amelioration of the disease  
CC or disorder such as inherited neurological and metabolic diseases e.g.  
CC lysosomal storage disease, Leech-Nyhan syndrome, amyloid polynuropathy,  
CC Alzheimer amyloid, Duchenne's muscular dystrophy, diseases of the blood  
CC e.g. sickle-cell anemia, clotting disorders and thalassemias, cystic  
CC fibrosis, diabetes, diseases associated with hormone deficiencies,  
CC retinoblastoma and various types of neoplastic cells which include  
CC tumours especially central nervous system tumours, neoplasms, carcinomas,  
CC sarcomas, leukemias and lymphoma. The present sequence represents the  
CC adeno-associated virus 2 (AAV-2) major coat protein VP1. AAV-2 sequences

CC are used in the construction of a chimeric vector.  
XX  
SQ Sequence 735 AA;

Query Match 85.4%; Score 2481.5; DB 22; Length 735;  
Best Local Similarity 83.3%; Pred. No. 9.8e-201;  
Matches 445; Conservative 37; Mismatches 51; Indels 1; Gaps 1;

QY 1 MASGGAPMADNNEGADGVGNASGNMHCDSWLGDRVITSTRTWALPTYNNHLKYQISS 60  
||:|||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
DB 203 MATSGAPMADNNEGADGVGNSSGNMHCDSWMDRVITSTRTWALPTYNNHLKYQISS 262  
  
QY 61 ASTGASNDNHFGYSTPWGYFDENRFCHFSPRDWQRLINNNWGFRPKRLNFKLFNIQVK 120  
||:|||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
DB 263 QS-GASNDNHFGYSTPWGYFDENRFCHFSPRDWQRLINNNWGFRPKRLNFKLFNIQVK 321  
  
QY 121 EVTTNDGVTTIANNLTSTVQVFSDEYQLPYVLGSAHQGLPFPADVFMIPQYGYLTIN 180  
|||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
DB 322 EVTQNDGTTTIANNLTSTVQVFTDSEYQLPYVLGSAHQGLPFPADVFMVPOYGYLTIN 381  
  
QY 181 NSQAVGRSSFYCLEYFSPQMLRTGNNFTFSYTEEVPFHSSYAHQSGLDRLMNPILIDQY 240  
|||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
DB 382 NSQAVGRSSFYCLEYFSPQMLRTGNNFTFSYTEEVPFHSSYAHQSGLDRLMNPILIDQY 441  
  
QY 241 LYYLNRTQNGSGAQNKDLLFSRGSFAGMSVQPKMWLPGPCYRQQRVSKTKTDNNNSNFT 300  
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|  
DB 442 LYYLSRTNTPSGTTTQSRLOFSQAGASDIRDQSRNWLPGPCYRQQRVSKTSADNNNSEYS 501  
  
QY 301 WTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMGVMIFGKESAGASNTALDNVMITD 360  
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|  
DB 502 WTGATKYHLNGRDSLVPNGPAMASHKDEEKFPPQSGVLI FGKQGESEKTNVDIEKVMITD 561  
  
QY 361 EEEIKATNPVATERFGTVAVNFQSSSTDPATGVDHAMGALPGMWQDRDYYLQGPWAKI 420  
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|  
DB 562 EEEIRTNPVATEQYGSVSTNLQGRNQAAATADVNTQGVLPGMWQDRDYYLQGPWAKI 621  
  
QY 421 PHTDGHFSPPLMGFGGLKNPPOILLIKNTPVPANPPAEFSATKFAFITQYSTGQVSVE 480  
|||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
DB 622 PHTDGHFSPPLMGFGGLKNPPOILLIKNTPVPANPSTTFSAAKFAFITQYSTGQVSVE 681  
  
QY 481 IEWELQKNSKRWNPVEQYTSNYAKSANVDFTVDNNGLYTEBRPIGTRYLTRPL 534  
|||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
DB 682 IEWELQKNSKRWNPVEIQYTSNYKSVNVDFTVDNNGVYSEBRPIGTRYLTRNL 735

RESULT 11  
AAM51508  
ID AAM51508 standard; protein; 735 AA.  
XX  
AC AAM51508;  
XX  
DT 02-JAN-2002 (first entry)  
XX  
DE Adeno-associated virus VP1 capsid protein.  
XX  
KW Adeno-associated virus; AAV; VP1; capsid; virus-like particle;  
KW nuclear localisation signal; VP3.  
XX  
OS Adeno associated virus.  
XX  
PN JP2001169777-A.  
XX  
PD 26-JUN-2001.  
XX  
PF 30-JUL-1999; 99JP-0249140.  
XX  
PR 30-JUL-1999; 99JP-0249140.  
XX  
PA (HAND/) HANDA H.  
XX  
DR WPI; 2001-599854/68.  
XX  
PT New virus-like particles from VP3 capsid protein of adeno-associated

PT virus, comprise a peptide containing a nucleus-shifting signal  
PT connected to its N-terminal -  
XX  
PS Disclosure; Page 10-13; 33pp; Japanese.  
XX

CC The present sequence is provided in a specification relating to a  
CC virus-like particle-forming protein, and to a peptide containing a  
CC nuclear-shifting signal at its N-terminus that can form virus-like  
CC particles by shifting to the nucleus of the animal cell in which it  
CC is expressed. The method is used for forming virus-like particles  
CC from the VP3 capsid protein of adeno-associated virus (AAV).  
CC  
XX

SQ Sequence 735 AA;  
  
Query Match 85.4%; Score 2481.5; DB 22; Length 735;  
Best Local Similarity 83.3%; Pred. No. 9.8e-201;  
Matches 445; Conservative 37; Mismatches 51; Indels 1; Gaps 1;

QY 1 MASGGAPMADNNEGADGVGNASGNMHCDSWLGDRVITSTRTWALPTYNNHLKYQISS 60  
||:|||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
DB 203 MATSGAPMADNNEGADGVGNSSGNMHCDSWMDRVITSTRTWALPTYNNHLKYQISS 262  
  
QY 61 ASTGASNDNHFGYSTPWGYFDENRFCHFSPRDWQRLINNNWGFRPKRLNFKLFNIQVK 120  
|||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
DB 263 QS-GASNDNHFGYSTPWGYFDENRFCHFSPRDWQRLINNNWGFRPKRLNFKLFNIQVK 321  
  
QY 121 EVTTNDGVTTIANNLTSTVQVFSDEYQLPYVLGSAHQGLPFPADVFMIPQYGYLTIN 180  
|||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
DB 322 EVTQNDGTTTIANNLTSTVQVFTDSEYQLPYVLGSAHQGLPFPADVFMVPOYGYLTIN 381  
  
QY 181 NSQAVGRSSFYCLEYFSPQMLRTGNNFTFSYTEEVPFHSSYAHQSGLDRLMNPILIDQY 240  
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|  
DB 382 NSQAVGRSSFYCLEYFSPQMLRTGNNFTFSYTEEVPFHSSYAHQSGLDRLMNPILIDQY 441  
  
QY 241 LYYLNRTQNGSGAQNKDLLFSRGSFAGMSVQPKMWLPGPCYRQQRVSKTKTDNNNSNFT 300  
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|  
DB 442 LYYLSRTNTPSGTTTQSRLOFSQAGASDIRDQSRNWLPGPCYRQQRVSKTSADNNNSEYS 501  
  
QY 301 WTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMGVMIFGKESAGASNTALDNVMITD 360  
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|  
DB 502 WTGATKYHLNGRDSLVPNGPAMASHKDEEKFPPQSGVLI FGKQGESEKTNVDIEKVMITD 561  
  
QY 361 EEEIKATNPVATERFGTVAVNFQSSSTDPATGVDHAMGALPGMWQDRDYYLQGPWAKI 420  
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|  
DB 562 EEEIRTNPVATEQYGSVSTNLQGRNQAAATADVNTQGVLPGMWQDRDYYLQGPWAKI 621  
  
QY 421 PHTDGHFSPPLMGFGGLKNPPOILLIKNTPVPANPPAEFSATKFAFITQYSTGQVSVE 480  
|||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
DB 622 PHTDGHFSPPLMGFGGLKNPPOILLIKNTPVPANPSTTFSAAKFAFITQYSTGQVSVE 681  
  
QY 481 IEWELQKNSKRWNPVEQYTSNYAKSANVDFTVDNNGLYTEBRPIGTRYLTRPL 534  
|||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
DB 682 IEWELQKNSKRWNPVEIQYTSNYKSVNVDFTVDNNGVYSEBRPIGTRYLTRNL 735

RESULT 12  
AAB59844  
ID AAB59844 standard; Protein; 735 AA.  
XX  
AC AAB59844;  
XX  
DT 28-MAR-2001 (first entry)  
XX  
DE AAV2 capsid protein VP1.  
XX  
DE AAV2 capsid protein VP1.  
XX  
KW AAV2; gene therapy; AAV viral vector; cystic fibrosis; cancer; AIDS;  
KW atherosclerosis; sickle cell anaemia; thalassaemia;  
KW blood clotting disorder; diabetes; capsid protein VP1.  
XX  
OS Adeno associated virus.  
XX  
PN US6156303-A.  
XX

PD 05-DEC-2000.  
XX  
XX 11-JUN-1997; 97US-0873168.  
PF  
XX 11-JUN-1997; 97US-0873168.  
PR  
XX (UNIW ) UNIV WASHINGTON.  
PA  
XX Russell DW, Rutledge EA;  
PI  
XX WPI; 2001-060164/07.  
DR  
XX  
XX Adeno-associated virus serotype 6 and viral vector derived from it for  
PT gene therapy of cystic fibrosis, cancer, acquired immunodeficiency  
PT syndrome, sickle cell anemia, thalassemia and diabetes  
XX  
XX  
PS Claim 7; Fig 2; 50pp; English.  
XX  
CC The present invention relates to adeno-associated virus serotypes. The  
CC present sequence is capsid protein VP1 of one such serotype (AAV2). AAV2  
CC can be used to construct AAV viral vectors for use in gene therapy for a  
CC range of disorders: cystic fibrosis, cancer, AIDS, atherosclerosis,  
CC sickle cell anaemia, thalassemia, blood clotting disorders and diabetes.  
CC The AAV viral vectors have increased transduction efficiency of a  
CC particular host cell as the AAV virion containing the AAV vector genome  
CC can be modified to express a capsid protein of an AAV serotype that  
CC transduces the selected host cell.  
XX  
SQ Sequence 735 AA;

Query Match 85.4%; Score 2481.5; DB 22; Length 735; \*  
Best Local Similarity 83.3%; Pred. No. 9.8e-201;  
Matches 445; Conservative 37; Mismatches 51; Indels 1; Gaps 1;

QY 1 MASGGGAPMADNNEGADGVGNASGNWCHDSTWLGDRVITTTSTRTWALPTYNHLYKQISS 60  
DB 203 MATGSGAPMADNNEGADGVGNSSGNWCHDSTWMDRVITTTSTRTWALPTYNHLYKQISS 262  
QY 61 ASTGASNDNHFGYSTPWCYFDFNRFCHFSPRDWQRLINNNWGFRPKRLNFKLFNIQVK 120  
DB 263 QS-GASNDNHFGYSTPWCYFDFNRFCHFSPRDWQRLINNNWGFRPKRLNFKLFNIQVK 321  
QY 121 EYTTNDGVTTIANNLTSTVQVFSDEYQLPYVLGSAHQCLPPFPADVEMIPOYGYLTIN 180  
DB 322 EYTONDGTITIANNLSTVQVFTDSEYQLPYVLGSAHQCLPPFPADVEMVPOYGYLTIN 381  
QY 181 NGSQAVGRSSFYCLEYFSPQMLRTGNNFTFSYTFEEVPHSSSYAHSQSLDRLMPLIDQY 240  
DB 382 NGSQAVGRSSFYCLEYFSPQMLRTGNNFTFSYTFEEVPHSSSYAHSQSLDRLMPLIDQY 441  
QY 241 LYTLNRTQNSGSAQNKDLFSRGSFAGMSVQPKNMLPGPCYRQORVSKTKTDNNNSNFT 300  
DB 442 LYTLNRTNTPSGTTTQSLRQFSQAGASDIRDQSRNMLPGPCYRQORVSKTSADNNSEYS 501  
QY 301 WTGASKYNLNGRESITINPGTAMASHKDEDEKFFPMGVMIFGKESAGASNTALDNVMTD 360  
DB 502 WTGATKYHLNGRDSLVPNGPAMASHKDEDEKFFPQSGVLIIFGKQSEKTNVDIEKVMITD 561  
QY 361 EEEIKATNPVATERFGTVAVNFQSSSTDPATGVDHAMGALPGMWQDRDVLVLOGPWAKI 420  
DB 562 EEEIRTNPVATEQYGSVSTNLQGRNRQAATADVNTQGVLPGMWQDRDVLVLOGPWAKI 621  
QY 421 PHTDGHFHPSPMLMGFGFLKNPPOILLIKNTPVPPANPFAFSATKFAFITQYSTGQVSVE 480  
DB 622 PHTDGHFHPSPMLMGFGFLKNPPOILLIKNTPVPPANPSTTFSAKFAFITQYSTGQVSVE 681  
QY 481 IEWELQKENSKRWNPEVQYTSNYSKASAVDFTVDNNGLYTEPRPIGTRYLTRPL 534  
DB 682 IEWELQKENSKRWNPEIQYTSNYSKASAVDFTVDNNGVSEPRPIGTRYLTRNL 735

RESULT 13  
AAM51510

ID AAM51510 standard; protein; 533 AA.  
XX  
XX AC AAM51510;  
XX  
XX DT 02-JAN-2002 (first entry)  
XX  
XX DE Adeno-associated virus VP3 capsid protein.  
XX  
XX KM Adeno-associated virus; AAV; capsid; virus-like particle;  
XX KM nuclear localisation signal; VP3.  
XX  
XX OS Adeno associated virus.  
XX  
XX PN JP2001169777-A.  
XX  
XX PD 26-JUN-2001.  
XX  
XX PF 30-JUL-1999; 99JP-0249140.  
XX  
XX PR 30-JUL-1999; 99JP-0249140.  
XX  
XX PA (HAND/) HANDA H.  
XX  
XX DR WPI; 2001-599854/68.  
XX  
XX PT New virus-like particles from VP3 capsid protein of adeno-associated  
PT virus, comprise a peptide containing a nucleus-shifting signal  
PT connected to its N-terminal -  
XX  
XX PS Disclosure; Page 17-19; 33pp; Japanese.  
XX  
XX CC The present sequence is provided in a specification relating to a  
CC virus-like particle-forming protein, and to a peptide containing a  
CC nuclear-shifting signal at its N-terminus that can form virus-like  
CC particles by shifting to the nucleus of the animal cell in which it  
CC is expressed. The method is used for forming virus-like particles  
CC from the VP3 capsid protein of adeno-associated virus (AAV).  
XX  
SQ Sequence 533 AA;

Query Match 85.3%; Score 2477.5; DB 22; Length 533;  
Best Local Similarity 83.3%; Pred. No. 1.3e-200;  
Matches 445; Conservative 36; Mismatches 52; Indels 1; Gaps 1;

QY 1 MASGGGAPMADNNEGADGVGNASGNWCHDSTWLGDRVITTTSTRTWALPTYNHLYKQISS 60  
DB 1 MATGSGAPMADNNEGADGVGNSSGNWCHDSTWMDRVITTTSTRTWALPTYNHLYKQISS 60  
QY 61 ASTGASNDNHFGYSTPWCYFDFNRFCHFSPRDWQRLINNNWGFRPKRLNFKLFNIQVK 120  
DB 61 QS-GASNDNHFGYSTPWCYFDFNRFCHFSPRDWQRLINNNWGFRPKRLNFKLFNIQVK 119  
QY 121 EYTTNDGVTTIANNLTSTVQVFSDEYQLPYVLGSAHQCLPPFPADVEMIPOYGYLTIN 180  
DB 120 EYTONDGTITIANNLSTVQVFTDSEYQLPYVLGSAHQCLPPFPADVEMVPOYGYLTIN 179  
QY 181 NGSQAVGRSSFYCLEYFSPQMLRTGNNFTFSYTFEEVPHSSSYAHSQSLDRLMPLIDQY 240  
DB 180 NGSQAVGRSSFYCLEYFSPQMLRTGNNFTFSYTFEEVPHSSSYAHSQSLDRLMPLIDQY 239  
QY 241 LYTLNRTQNSGSAQNKDLFSRGSFAGMSVQPKNMLPGPCYRQORVSKTKTDNNNSNFT 300  
DB 240 LYTLNRTNTPSGTTTQSLRQFSQAGASDIRDQSRNMLPGPCYRQORVSKTSADNNSEYS 299  
QY 301 WTGASKYNLNGRESITINPGTAMASHKDEDEKFFPMGVMIFGKESAGASNTALDNVMTD 360  
DB 300 WTGATKYHLNGRDSLVPNGPAMASHKDEDEKFFPQSGVLIIFGKQSEKTNVDIEKVMITD 359  
QY 361 EEEIKATNPVATERFGTVAVNFQSSSTDPATGVDHAMGALPGMWQDRDVLVLOGPWAKI 420  
DB 360 EEEIRTNPVATEQYGSVSTNLQGRNRQAATADVNTQGVLPGMWQDRDVLVLOGPWAKI 419  
QY 421 PHTDGHFHPSPMLMGFGFLKNPPOILLIKNTPVPPANPFAFSATKFAFITQYSTGQVSVE 480





CC provides recombinant vectors and viral particles based on AAV4 that  
 CC may be useful for transducing erythroid progenitor cells for the  
 CC treatment of e.g. cancer and genetic diseases which can be corrected  
 CC by bone marrow transplants using matched donors. A claimed method  
 CC of delivering a nucleic acid to a subject (including a subject with  
 CC antibodies to AAV2) comprises administering to a cell from the  
 CC subject an AAV4 particle comprising the nucleic acid inserted  
 CC between a pair of AAV inverted terminal repeats (see AAV21651 and  
 CC AAV21659), and returning the cell to the subject. An antibody  
 CC against AAV4 VP2 can be used to determine the suitability of an  
 CC AAV4 vector for administration to a subject. AAV4 VP2 can also be  
 CC used to detect the presence of AAV4-specific antibodies in a sample.

XX Sequence 544 AA;

Query Match 58.2%; Score 1690.5; DB 19; Length 544;  
 Best Local Similarity 59.4%; Pred. No. 4.3e-134;  
 Matches 325; Conservative 65; Mismatches 132; Indels 25; Gaps 9;

OY 2 ASGGAPMADNNEGADGVNAGSNWCHDSTWLGDRVITTTSTRTWALPTYNHLYKQISSA 61  
 Db 9 AAAGGAAY-EGGAGADGVNAGSNWCHDSTWSEGHVTTTSTRTWLPTYNHLYKRLGE- 66  
 OY 62 STGASNDNHFGYSTPWGYPDFNRFCHSPRDWQRLINNNGFRPKRLNFKLFNIQVKE 121  
 Db 67 ---SLQSNYNGFSTPWGYPDFNRFCHSPRDWQRLINNNGFRPKRLNFKLFNIQVKE 123  
 OY 122 VTNDGVTIANNLSTVQVFSDEYQPYVLGSAHQGLPPFPADVFMIPQYGY--LT 178  
 Db 124 VTTNGETVANNLTSTVQIFADSSYELPYMDAGQEGSLPPFPADVFMIPQYGYCGLVT 183  
 OY 179 LNNGSAVGRSSFCLEYPFSQMLRTGNNEFTSYTFEBVPFHSVYASQSLDRLMNPID 238  
 Db 184 GNTSQQGTDRNAFYCLEYFSPQMLRTGNNEFTSYTFEBVPFHSVYASQSLDRLMNPID 243  
 OY 239 QYLYLNRTON----QSGSAQNKDLLFSRGSFAGMSVQPKNWLPGPCYRQQRVSKTKTDN 294  
 Db 244 QYLMGLQSTTGTTLNAGTATTN--FTKLAPTFNFSNFKKNWLPGPSIKOQGFSKTA--N 298  
 OY 295 NNSNFTWTGAS---KY---NLNGRESINDGTAMASHKDEDEKFFPMGVMIFGKESAG 347  
 Db 299 QNYKIPATGSDSLIKYETHSTLDGRWSALTPGPPMATAGPADSK-FSNSQLIFAGPKONG 357  
 OY 348 ASNTALDNVMTDDEEIKATNPVATERFGTAVAVNFQSSSTDPATGDVHAMGALPGMWQD 407  
 Db 358 NTATVPGLTIFTSEELATNATDITDMWGNLPGDQNSNLPVIVRLTALGAVPGMWQD 417  
 OY 408 RDVYLQGPPIWAKIPHTDGFHPSPLMGFGFKNPPOILIKNTPVPANPAPAEFSATKFA 467  
 Db 418 RDIYVQGPPIWAKIPHTDGFHPSPLMGFGFKNPPQIFIKNTPVPANPATFSSSTPVNS 477  
 OY 468 FITQYSTGVSVIEIEMELQKENSKRWNPEVOYTSNYAKSANVDFTVNNGLYTEBRPIGT 527  
 Db 478 FITQYSTGVSVQIDWEIQKERSKRWNPEVOFTSNYGQNSLIMAPDAAGKYTEBRAIGT 537  
 OY 528 RYLTRPL 534  
 Db 538 RYLTHHL 544

Search completed: July 17, 2003, 18:31:21  
 Job time : 38.1429 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 2003, 18:29:05 ; Search time 13.1429 Seconds  
(without alignments)  
1195.466 Million cell updates/sec

Title: US-09-807-802A-17  
Perfect score: 2906  
Sequence: 1 MASGGGAPMADNNEGADGVG.....NNGLYTEPRPIGTRYLTRPL 534

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1690.5	58.2	544	4	US-09-532-594B-18 Sequence 18, Appl
2	1690.5	58.2	598	4	US-09-532-594B-16 Sequence 16, Appl
3	1690.5	58.2	734	4	US-09-532-594B-4 Sequence 4, Appl
4	479.5	16.5	543	4	US-08-856-841-22 Sequence 22, Appl
5	326	11.2	415	4	US-08-856-841-20 Sequence 20, Appl
6	318	10.9	395	4	US-08-856-841-13 Sequence 13, Appl
7	318	10.9	486	4	US-08-856-841-19 Sequence 19, Appl
8	317.5	10.9	500	4	US-08-856-841-16 Sequence 16, Appl
9	317.5	10.9	501	4	US-08-856-841-18 Sequence 18, Appl
10	304.5	10.5	264	4	US-08-856-841-14 Sequence 14, Appl
11	303	10.4	398	4	US-08-856-841-21 Sequence 21, Appl
12	209.5	7.2	579	6	5223424-13 Patent No. 5223424
13	183	6.3	584	4	US-09-022-949-2 Sequence 2, Appl
14	171	5.9	387	4	US-08-856-841-17 Sequence 17, Appl
15	111	3.8	655	1	US-08-469-202-27 Sequence 27, Appl
16	111	3.8	655	2	US-08-469-202-27 Sequence 27, Appl
17	109	3.8	655	1	US-08-484-434C-34 Sequence 34, Appl
18	109	3.8	655	2	US-08-484-434C-34 Sequence 34, Appl
19	109	3.8	655	1	US-08-485-568A-4 Sequence 4, Appl
20	109	3.8	655	2	US-08-485-568A-4 Sequence 4, Appl
21	109	3.8	655	1	US-08-590-554A-4 Sequence 4, Appl
22	108	3.7	624	3	US-09-184-223-4 Sequence 4, Appl
23	108	3.7	624	3	US-09-184-223-4 Sequence 4, Appl
24	108	3.7	655	1	PCT-US93-12682-6 Sequence 6, Appl
25	104.5	3.6	717	4	US-08-947-965-78 Sequence 78, Appl
26	104	3.6	717	4	US-08-947-965-78 Sequence 78, Appl
27	104	3.5	1013	4	US-08-484-434C-35 Sequence 35, Appl
					US-09-626-589-1 Sequence 1, Appl
					US-09-415-522-8 Sequence 8, Appl
					US-09-134-001C-3159 Sequence 3159, Ap

28	99	3.4	824	4	US-09-626-589-3	Sequence 3, Appl
29	98.5	3.4	997	1	US-08-232-540-1	Sequence 1, Appl
30	98.5	3.4	997	1	US-08-428-949A-1	Sequence 1, Appl
31	98.5	3.4	997	1	US-08-428-948A-1	Sequence 1, Appl
32	98.5	3.4	997	2	US-08-428-946-1	Sequence 1, Appl
33	98.5	3.4	997	5	PCT-US95-04656-1	Sequence 1, Appl
34	98.5	3.4	998	1	US-08-233-008A-6	Sequence 6, Appl
35	98.5	3.4	1000	4	US-09-193-562D-30	Sequence 30, Appl
36	98.5	3.4	1021	1	US-08-233-008A-2	Sequence 2, Appl
37	98	3.4	1178	6	5254799-5 Patent No. 5254799	Sequence 7, Appl
38	97	3.3	540	4	US-08-687-580B-7	Sequence 2, Appl
39	97	3.3	659	4	US-09-626-589-2	Sequence 18, Appl
40	96.5	3.3	1651	4	US-09-540-245A-18	Sequence 1, Appl
41	96.5	3.3	2237	1	US-08-354-973-1	Patent No. 5169629
42	96	3.3	1177	6	5169629-2	Sequence 2, Appl
43	96	3.3	1394	4	US-08-296-791-2	Sequence 2, Appl
44	96	3.3	1394	5	PCT-US95-10661A-2	Sequence 77, Appl
45	95	3.3	680	3	US-08-947-965-77	

ALIGNMENTS

RESULT 1  
US-09-532-594B-18  
; Sequence 18, Application US/09532594B  
; Patent No. 6468524  
; GENERAL INFORMATION:  
; APPLICANT: Chorini, John A.  
; APPLICANT: Kotin, Robert M.  
; APPLICANT: Safer, Brian  
; APPLICANT: Davidson, Beverly  
; TITLE OF INVENTION: AAV4 VECTOR AND USES THEREOF  
; FILE REFERENCE: 14014.0252U2  
; CURRENT APPLICATION NUMBER: US/09/532,594B  
; CURRENT FILING DATE: 2000-03-22  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 18  
; LENGTH: 544  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence; No. 6468524e =  
; NAME/KEY: misc feature  
; OTHER INFORMATION: AAV4 capsid protein VP3  
US-09-532-594B-18

QY	2	ASGGAPMADNNEGADGVGNAGNWHCDSTWLGDRTVITSTRWALLPTYNHLYKQISSA	61
DB	9	AAAGGAIV-EGGQADGVGNAGNWHCDSTWSEGHVTTSTRTWLPTYNHLYKRLGE-	66
QY	62	STGASNDNHYPGYSTPWGYPFDRNRFHCHFSPRDWQRLINNNGFRPKRLNFKLFNIQVKE	121
DB	67	---SLOSNTYNGFSTPWGYPFDRNRFHCHFSPRDWQRLINNNGFRPKRLNFKLFNIQVKE	123
QY	122	VTTNDGVTIANMLTSTVQVPSDSEYQLPYVLGSAHQGCLPPPADVFMIPQYGY--LT	178
DB	124	VTTNSGRTTAVNNLTSTVQIFADSSYELPYVMDAGQEGSLPPFPNDVFMVPOYGCGLVT	183
QY	179	LNNGSQAVGRSSFYCLEYFPQWMLRTGNFTFSYTFEEVPEHSSYASQSLDRLMPLID	238
DB	184	GNTSOQOTDRNAFYCLYFPQWMLRTGNFTFSYTFEEVPEHSSYASQSLDRLMPLID	243
QY	239	QYLYVLRTON---QSGSAONKDLLFSRGSFAGMSVQPKWLLPGPCYRQORVSKTKTDN	294
DB	244	QYLMGLOSTTGTTLNAGTATNN---FTKLRTPTFSNFKKNWLLPGPSIKQOGFSKTA--N	298

[illegible]

## RESULT 2

```

US-09-532-594B-16
; Sequence 16, Application US/09532594B
; Patent No. 6468524
; GENERAL INFORMATION:
; APPLICANT: Chorini, John A.
; APPLICANT: Kotin, Robert M.
; APPLICANT: Safer, Brian
; APPLICANT: Davidson, Beverly
; TITLE OF INVENTION: AAV4 VECTOR AND USES THEREOF
; FILE REFERENCE: 14014.0252U2
; CURRENT APPLICATION NUMBER: US/09/532,594B
; CURRENT FILING DATE: 2000-03-22
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 598
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence; No. 6468524e =
; OTHER INFORMATION: synthetic construct
; NAME/KEY: misc_feature
; OTHER INFORMATION: AAV4 capsid protein VP2
US-09-532-594B-16

```

[illegible][illegible]

### RESULT 3

```

US-09-532-594B-4
; Sequence 4, Application US/09532594B
; Patent No. 6468524
; GENERAL INFORMATION:
; APPLICANT: Chorini, John A.
; APPLICANT: Kotin, Robert M.
; APPLICANT: Safer, Brian
; APPLICANT: Davidson, Beverly
; TITLE OF INVENTION: AAV4 VECTOR AND USES THEREOF
; FILE REFERENCE: 14014.0252U2
; CURRENT APPLICATION NUMBER: US/09/532,594B
; CURRENT FILING DATE: 2000-03-22
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for windows Version 4.0
; SEQ ID NO 4
; LENGTH: 734
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence; No. 6468524e =
; OTHER INFORMATION: synthetic construct
; NAME/KEY: misc_feature
; OTHER INFORMATION: AAV4 capsid protein VP1
US-09-532-594B-4

```

[illegible]

QY 348 ASNTALDNVMTDDEEIKATNPVATERFGIVA VNFQSSSTDPATGVDVHAMGALPGMWOD 407  
Db 548 NTATVPGTILFTSEELATNATDTDMWGNLPGGDQSNLSNLPVTDRLTALGAVPGMWON 607  
QY 408 RDVYLQGPIMAKI PHTDGHFHPSPLMGFGIKNPPQILIKNTPVANPAPAEFSATKFA 467  
Db 608 RDIYQGPIMAKI PHTDGHFHPSPLMGFGIKNPPQILIKNTPVANPAPAEFSATKFA 667  
QY 468 FITQYSTGVSEIEMELQKENS KRWNPVQYTSNYAKSANVDFVNNGLYTEPRPIGT 527  
Db 668 FITQYSTGVSEIEMELQKENS KRWNPVQYTSNYAKSANVDFVNNGLYTEPRPIGT 727  
QY 528 RYLTRPL 534  
Db 728 RYLTRPL 734

RESULT 4

US-08-856-841-22  
; Sequence 22, Application US/08856841  
; Patent No. 6274307  
; GENERAL INFORMATION:  
; APPLICANT: ERWIN SOUTSCHEK  
; APPLICANT: MANFRED MOTZ  
; TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES  
; TITLE OF INVENTION: OR POLYPEPTIDES FROM THE PARVOVIRUS B19  
; NUMBER OF SEQUENCES: 28  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROOKS HAIDT HAFNER & DELAHUNTY  
; STREET: 99 PARK AVENUE  
; CITY: NEW YORK  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10016  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" FLOPPY DISC  
; COMPUTER: AT&T - IBM COMPATIBLE  
; OPERATING SYSTEM: MS-DOS Version 6.2  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/856,841  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/214,658  
; FILING DATE: 16-MARCH-1994  
; APPLICATION NUMBER: US 07/917,096  
; FILING DATE: 4-AUGUST-1992  
; APPLICATION NUMBER: PCT/DE91/00106  
; FILING DATE: 8-FEBRUARY-1991  
; APPLICATION NUMBER: DE40038262  
; FILING DATE: 8-FEBRUARY-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: ROBINSON, WILLIAM R.  
; REGISTRATION NUMBER: 27,224  
; REFERENCE/DOCKET NUMBER: LKR-9222-A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 697-3355  
; TELEFAX: (212) 557-5635  
; TELEX: NONE  
; INFORMATION FOR SEQ ID NO: 22:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 543  
; TYPE: AMINO ACID  
; TOPOLOGY: LINEAR  
; MOLECULE TYPE:  
; DESCRIPTION: PEPTIDE  
; HYPOTHETICAL: N/A  
; ANTI-SENSE: N/A  
; FRAGMENT TYPE: INTERNAL  
; ORIGINAL SOURCE: SERUM FROM PATIENT WITH ACUTE  
; IMMEDIATE SOURCE: GENETICALLY ENGINEERED PEPTIDE

; POSITION IN GENOME: N/A  
; FEATURE:  
; NAME/KEY: N/A  
; LOCATION: N/A  
; IDENTIFICATION METHOD: amino acid analysis and  
; IDENTIFICATION METHOD: mass spectrometry  
; OTHER INFORMATION:  
; PUBLICATION INFORMATION:  
; AUTHORS: COSSART, Y.E.  
; AUTHORS: FIELD, A.M.  
; AUTHORS: CANT, B.  
; AUTHORS: WIDDOWS, D.  
; TITLE: PARVOVIRUS-LIKE PARTICLES IN HUMAN SERA  
; JOURNAL: LANCET  
; VOLUME: I  
; ISSUE:  
; PAGES: 72 - 73  
; DATE: 1975  
; DOCUMENT NUMBER:  
; FILING DATE:  
; PUBLICATION DATE:  
; RELEVANT RESIDUES IN SEQ ID NO: 22:  
US-08-856-841-22

Query Match 16.5%; Score 479.5; DB 4; Length 543;  
Best Local Similarity 27.1%; Pred. No. 1.9e-35;  
Matches 141; Conservative 79; Mismatches 226; Indels 75; Gaps 16;

QY 10 ADNNEGADGVGNAS--GNWHCDSTWLGDRVITSTRITWALPTYNHLYKOISSASTGASN 67  
Db 16 AEAAGTACGGGSGNSVKSMWSEGAITSANVTCTFSRQFLIPYDEHHYKVFSPAASCHN 75  
QY 68 D-----NHFGYSTPGYFDNRFHCHFSRPRDQRLNNWGFPRKRLNFKLFNI 117  
Db 76 ASGKAQVCTISPIMGYSTPWRYLDFNALNLFSSPLEFQHLIENYGSIAADALTVTISEI 135  
QY 118 QVKEVT--TNDGVTIANNLSTQVSESEYQLPYVLGSAHQCLPPFPADVFMIPQYG 175  
Db 136 AVKDVTDKGGV-QVTDSTGRCLMLVDHEKYRYVLGGQDTLAPLPIWVFPQYA 194  
QY 176 YLTLLN-GSQAVG-----RSSFYCLEYFSPQMLRTGNFTFSYTFEEVPFHSSYAH 225  
Db 195 YLTGVDVNTQGISGDSKSLASEESA FYVLEHSSFQLGTGTASMSYKFPVPPEBLEGC 254  
QY 226 SQSLDRMLNPLIDQYLYLNRTQNSGSAQNKDLFSRGSFAGMSVQPKWLPGPCYRQ 285  
Db 255 SQHFYEMVNP--YGSRLGVDPDLGGDPKFRSL----THEDHAIQPNFMFGPLVNSV 306  
QY 286 RVSKTYTDNNNSNFTWTGASKYNLNGRESIINPG-TAMASHKDEDEKFFPMGVMIFGE 344  
Db 307 STKEGDSNTGAKALTGLSTGTSQNTRISLRPGPVSQPYHHWDTDKYVTGINAISHGT 366  
QY 345 SAGASNTALDNV-----MITDEEIKATNPVATERFGIVA VNFQSSSTDPATGVDHAM 397  
Db 367 TYG---NAEDKEYQGVGRFPNEKEQLKQLGLNMHTY-----FPNKGTOQYTDQIE-R 416  
QY 398 GALPGWVQDRDVYLQGPIMAKI PHTDGHFHPSPLMGFGIKNPPQILIKNTPVANP 456  
Db 417 PLWGSVWNRALHYESQLSKIPNLDDSFKTQFAALGSGWGLHQPPQI----- 465  
QY 457 PAEFSATKFAFITQYSTGVSEIEMEL-QKENS KRWNP 496  
Db 466 -----FLKQYAVGIMVTMTFKLGRKATGRNRPQ 495

RESULT 5

US-08-856-841-20  
; Sequence 20, Application US/08856841  
; Patent No. 6274307  
; GENERAL INFORMATION:  
; APPLICANT: ERWIN SOUTSCHEK  
; APPLICANT: MANFRED MOTZ  
; TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES





TELEFAX: (212) 557-5635  
TELEX: NONE  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 395  
TYPE: AMINO ACID  
TOPOLOGY: LINEAR  
MOLECULE TYPE:  
DESCRIPTION: PEPTIDE  
HYPOTHETICAL: N/A  
ANTI-SENSE: N/A  
FRAGMENT TYPE: INTERNAL  
ORIGINAL SOURCE: SERUM FROM PATIENT WITH ACUTE  
INFECTION (ERYTHEMA INFECTIONOSUM)  
IMMEDIATE SOURCE: GENETICALLY ENGINEERED PEPTIDE  
POSITION IN GENOME: N/A  
FEATURE:  
NAME/KEY: N/A  
LOCATION: N/A  
IDENTIFICATION METHOD: amino acid analysis and  
IDENTIFICATION METHOD: mass spectrometry  
OTHER INFORMATION:  
PUBLICATION INFORMATION:  
AUTHORS: COSSART, Y.E.  
AUTHORS: FIELD, A.M.  
AUTHORS: CANT, B.  
AUTHORS: WIDDOWS, D.  
TITLE: PARVOVIRUS-LIKE PARTICLES IN HUMAN SERA  
JOURNAL: LANCET  
VOLUME: 1  
ISSUE:  
PAGES: 72 - 73  
DATE: 1975  
DOCUMENT NUMBER:  
FILING DATE:  
PUBLICATION DATE:  
RELEVANT RESIDUES IN SEQ ID NO: 13:  
PUBLICATION INFORMATION:  
AUTHORS: MANIATIS, T.  
AUTHORS: FRITSCH, E.F.  
AUTHORS: SAMBROOK, J.  
TITLE: MOLECULAR CLONING  
JOURNAL: COLD SPRING HARBOR, NY  
VOLUME:  
ISSUE:  
PAGES:  
DATE: 1982  
DOCUMENT NUMBER:  
FILING DATE:  
PUBLICATION DATE:  
RELEVANT RESIDUES IN SEQ ID NO: 13:  
PUBLICATION INFORMATION:  
AUTHORS: SMITH, D.B.  
AUTHORS: JOHNSON, K.S.  
TITLE: SINGLE STEP PURIFICATION OF POLYPEPTIDES  
TITLE: EXPRESSED IN ESCHERICHIA COLI AS FUSIONS WITH  
TITLE: GLUTATHIONE S. TRANSFERASE  
JOURNAL: GENE  
VOLUME:  
ISSUE: 67  
PAGES: 31 - 40  
DATE: 1988  
DOCUMENT NUMBER:  
FILING DATE:  
PUBLICATION DATE:  
RELEVANT RESIDUES IN SEQ ID NO: 13:  
US-08-856-841-13  
Query Match 10.9%; Score 318; DB 4; Length 395;  
Best local Similarity 25.4%; Pred. No. 7e-21;  
Matches 99; Conservative 61; Mismatches 167; Indels 62; Gaps 13;  
QY 128 VTTIANNTSTVQVSDSEYQLPYVLGSAHQCLPPFPADVFMPQYGYLTINN-GSQAV 186

Db 1 MTMTIPSLHACMLV--DHEYKXYVVLGGQODTLAPELPIMVYFPQYAYLTVDGVDNTQGI 58  
QY 187 G-----RSSFYCLEYFSPQMLRTGNNTFSYTFEEVPHSSYAHQSOLDRLMPLI 237  
Db 59 SGDSKILASEESAFAFVLEHSSFQILGTGTASMSYKFPVPPENLEGCSQHFYEMYNPL 117  
QY 238 DQYLYYINRTQNSGSAQNKDLFSRGSFAGMSVQPKNWLPGPCYRQQRVSKTKTDNNNS 297  
Db 118 --YGSRLGVPTLGGDPKFRSL-----THEDHAIQPNFMGPLVNSVSTKEGSSNTGA 170  
QY 298 NFWTWSKYNLNGRESIINPG-TAMASHKDEDEKFFPMGVMIFGESAGASNTALDNV 356  
Db 171 GKALTGLSTGTSQNTIRISLRPGVSPYHMDTDKYVTGAINAISHGQTTYG--NAEDKE 227  
QY 357 -----MITDEEIKATNPVATERFGTVAVNFQSSSTDPAIGDVHAMGALPGMWQDRD 409  
Db 228 YQGVGRFPNEKEQLKQLQGLNHTY-----FPNKGTOYTDQIE-RPLMVGSVMNRA 280  
QY 410 VYLQGPWAKIPHTDGHFHP-S-PLMGFGKLNPPQILLINTPVPANPPAEFSATKFAF 468  
Db 281 LHYESQMSKIPNLDSEFKTQPAALGGWGLHQPPOI-----F 318  
QY 469 ITQYSTQGVSEIEMEL-QKENSRRNPE 496  
Db 319 LKQYAVGIMTVTMTFKLGPRAKATGRWNPQ 347  
RESULT 7  
US-08-856-841-19  
Sequence 19, Application US/08856841  
Patent No. 6274307  
GENERAL INFORMATION:  
APPLICANT: ERWIN SOUTSCHEK  
APPLICANT: MANFRED MOTZ  
TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES  
TITLE OF INVENTION: OR POLYPEPTIDES FROM THE PARVOVIRUS B19  
NUMBER OF SEQUENCES: 28  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROOKS HAIDT HAFNER & DELAHUNTY  
STREET: 99 PARK AVENUE  
CITY: NEW YORK  
STATE: NY  
COUNTRY: USA  
ZIP: 10016  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" FLOPPY DISC  
COMPUTER: AT&T - IBM COMPATIBLE  
OPERATING SYSTEM: MS-DOS Version 6.2  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/856,841  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/214,658  
FILING DATE: 16-MARCH-1994  
APPLICATION NUMBER: US 07/917,096  
FILING DATE: 4-AUGUST-1992  
APPLICATION NUMBER: PCT/DE91/00106  
FILING DATE: 8-FEBRUARY-1991  
APPLICATION NUMBER: DE40038262  
FILING DATE: 8-FEBRUARY-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: ROBINSON, WILLIAM R.  
REGISTRATION NUMBER: 27,224  
REFERENCE/DOCKET NUMBER: LKR-9222-A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 697-3355  
TELEFAX: (212) 557-5635  
TELEX: NONE  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:







DB 422 YLTVGDVNTQGISGDSKSLASESAFYVLEHSSFQLLTGTGTASMSYKPPVPENLEG 481  
QY 226 SOSLDRLMPL 236  
DB 482 SQHFYEMYNPL 492

## RESULT 10

US-08-856-841-14  
; Sequence 14, Application US/08856841  
; Patent No. 6274307  
; GENERAL INFORMATION:  
; APPLICANT: ERWIN SOUTSCHEK  
; APPLICANT: MANFRED MOTZ  
; TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES  
; TITLE OF INVENTION: OR POLYPEPTIDES FROM THE PARVOVIRUS B19  
; NUMBER OF SEQUENCES: 28  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROOKS HAIDT HAFNER & DELAHUNTY  
; STREET: 99 PARK AVENUE  
; CITY: NEW YORK  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10016  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" FLOPPY DISC  
; COMPUTER: AT&T - IBM COMPATIBLE  
; OPERATING SYSTEM: MS-DOS Version 6.2  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/856,841  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/214,658  
; FILING DATE: 16-MARCH-1994  
; APPLICATION NUMBER: US 07/917,096  
; FILING DATE: 4-AUGUST-1992  
; APPLICATION NUMBER: PCT/DE91/00106  
; FILING DATE: 8-FEBRUARY-1991  
; APPLICATION NUMBER: DE40038262  
; FILING DATE: 8-FEBRUARY-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: ROBINSON, WILLIAM R.  
; REGISTRATION NUMBER: 27,224  
; REFERENCE/DOCKET NUMBER: LKR-9222-A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 697-3355  
; TELEFAX: (212) 557-5635  
; TELEX: NONE  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 264  
; TYPE: AMINO ACID  
; TOPOLOGY: LINEAR  
; MOLECULE TYPE: PEPTIDE  
; DESCRIPTION: N/A  
; HYPOTHETICAL: N/A  
; ANTI-SENSE: N/A  
; FRAGMENT TYPE: INTERNAL  
; ORIGINAL SOURCE: SERUM FROM PATIENT WITH ACUTE  
; IMMEDIATE SOURCE: GENETICALLY ENGINEERED PEPTIDE  
; POSITION IN GENOME: N/A  
; FEATURE:  
; NAME/KEY: N/A  
; LOCATION: N/A  
; IDENTIFICATION METHOD: amino acid analysis and  
; IDENTIFICATION METHOD: mass spectrometry  
; OTHER INFORMATION:  
; PUBLICATION INFORMATION:  
; AUTHORS: COSSART, Y.B.  
; AUTHORS: FIELD, A.M.

AUTHORS: CANT, B.  
AUTHORS: WIDOWS, D.  
TITLE: PARVOVIRUS-LIKE PARTICLES IN HUMAN SERA  
JOURNAL: LANCET  
VOLUME: 1  
ISSUE:  
PAGES: 72 - 73  
DATE: 1975  
DOCUMENT NUMBER:  
FILING DATE:  
PUBLICATION DATE:  
RELEVANT RESIDUES IN SEQ ID NO: 14:  
PUBLICATION INFORMATION:  
AUTHORS: MANIATIS, T.  
AUTHORS: FRITSCH, E.F.  
AUTHORS: SAMBROOK, J.  
TITLE: MOLECULAR CLONING  
JOURNAL: COLD SPRING HARBOR, NY  
VOLUME:  
ISSUE:  
PAGES:  
DATE: 1982  
DOCUMENT NUMBER:  
FILING DATE:  
PUBLICATION DATE:  
RELEVANT RESIDUES IN SEQ ID NO: 14:  
PUBLICATION INFORMATION:  
AUTHORS: SMITH, D.B.  
AUTHORS: JOHNSON, K.S.  
TITLE: SINGLE STEP PURIFICATION OF POLYPEPTIDES  
TITLE: EXPRESSED IN ESCHERICHIA COLI AS FUSIONS WITH  
TITLE: GLUTATHIONE S. TRANSFERASE  
JOURNAL: GENE  
VOLUME:  
ISSUE: 67  
PAGES: 31 - 40  
DATE: 1988  
DOCUMENT NUMBER:  
FILING DATE:  
PUBLICATION DATE:  
RELEVANT RESIDUES IN SEQ ID NO: 14:  
US-08-856-841-14

Query Match 10.5%; Score 304.5; DB 4; Length 264;  
Best Local Similarity 32.5%; Pred. No. 6.5e-20;  
Matches 76; Conservative 30; Mismatches 105; Indels 23; Gaps 5;

QY 7 APMADNNEGADGVGNASGNWCHDSTWLGDRTVITSTRITWALPTYNNHLYKQISSASTGAS 66  
DB 9 AAEASTGAGGGSNSVKSMWSEGAFTSANSVTCTFSRQFLIPDEHHYKVFSPAASSCH 68  
QY 67 ND-----NHFGYSTPWGTFDENRFHCHFSPRDWORLNNWGFPRKRLNFKLFN 116  
DB 69 NASGKAQVCTISPIMGYSTPWRYPDENALNLFSPLEFQHLIENYGSIAPDALTVTISE 128  
QY 117 IQVKEVT--TNDGVTTIANNTSTVQVFSDEYQLPYVLGSAHQGLPPEPADVFEMIPQY 174  
DB 129 IAVKDVTDKTTGGV-QVTDSTGRCLMLVDHXYKYPYVLGGGQDTLAPELPIWVYFPPQY 187  
QY 175 GYLTLNN-GSQAVG-----RSSFYCLEYFSPQMLRTGNFTFSYTEEVP 218  
DB 188 AYLTVGDVNTQGISGDSKSLASESAFYVLEHSSFQLLTGTGTASMSYKPPVP 241

## RESULT 11

US-08-856-841-21  
; Sequence 21, Application US/08856841  
; Patent No. 6274307  
; GENERAL INFORMATION:  
; APPLICANT: ERWIN SOUTSCHEK  
; APPLICANT: MANFRED MOTZ  
; TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES  
; TITLE OF INVENTION: OR POLYPEPTIDES FROM THE PARVOVIRUS B19

```

: NUMBER OF SEQUENCES: 28
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: BROOKS HAIDT HAFNER & DELAHUNTY
: STREET: 99 PARK AVENUE
: CITY: NEW YORK
: STATE: NY
: COUNTRY: USA
: ZIP: 10016
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5" FLOPPY DISC
: COMPUTER: AT&T - IBM COMPATIBLE
: OPERATING SYSTEM: MS-DOS Version 6.2
: SOFTWARE: ASCII
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/856,841
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/08/214,658
: FILING DATE: 16-MARCH-1994
: APPLICATION NUMBER: US 07/917,096
: FILING DATE: 4-AUGUST-1992
: APPLICATION NUMBER: PCT/DE91/00106
: FILING DATE: 8-FEBRUARY-1991
: APPLICATION NUMBER: DE40038262
: FILING DATE: 8-FEBRUARY-1990
: ATTORNEY/AGENT INFORMATION:
: NAME: ROBINSON, WILLIAM R.
: REGISTRATION NUMBER: 27,224
: REFERENCE/DOCKET NUMBER: LKR-9222-A
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 697-3355
: TELEFAX: (212) 557-5635
: TELEX: NONE
: INFORMATION FOR SEQ ID NO: 21:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 398
: TYPE: AMINO ACID
: TOPOLOGY: LINEAR
: MOLECULE TYPE:
: DESCRIPTION: PEPTIDE
: HYPOTHETICAL: N/A
: ANTI-SENSE: N/A
: FRAGMENT TYPE: INTERNAL
: ORIGINAL SOURCE: SERUM FROM PATIENT WITH ACUTE
: IMMEDIATE SOURCE: GENETICALLY ENGINEERED PEPTIDE
: POSITION IN GENOME: N/A
: FEATURE:
: NAME/KEY: N/A
: LOCATION: N/A
: IDENTIFICATION METHOD: amino acid analysis and
: IDENTIFICATION METHOD: mass spectrometry
: OTHER INFORMATION:
: PUBLICATION INFORMATION:
: AUTHORS: COSSART, Y.E.
: AUTHORS: FIELD, A.M.
: AUTHORS: CANT, B.
: AUTHORS: WIDDOWS, D.
: TITLE: PARVOVIRUS-LIKE PARTICLES IN HUMAN SERA
: JOURNAL: LANCET
: VOLUME: 1
: ISSUE:
: PAGES: 72 - 73
: DATE: 1975
: DOCUMENT NUMBER:
: FILING DATE:
: PUBLICATION DATE:
: RELEVANT RESIDUES IN SEQ ID NO: 21:
: US-08-856-841-21

```

```
QY      144 DSEYLPPVVLGSAHQGLPFPFADVMIPQYGTLTNN-GSQA-V-----RSSFYC   193
          |||:|||||
Db       4 DHEKYRYPVVGOGDTLAPELPIWVFPPQAYALTVDVNTQGISGDSKLAASEEAFIV    63
QY      194 LEYFPSOMLRITGNFTFSYTFEVEVPFHSSYAHSQSIDRLMNP.LIDQLYYLNRTQNOSGS  253
          ||:::|||||
Db        64 LHHSSFOLGTGGTA-TWSYKFPPVPPENLEGCSDHFEMYNPL---YGRLLGVPTLTGCD  120
QY      254 AQNKDLLFRSGSPAGMSVQPKNWLPGPCYRQRVS-KTKTDNNNSNFMTWTGA-SKYNLNGRE  313
          :.:|:|:|
Db      121 PKFRSL-THEDHAIQPNFMFGPLNVSVSTKEGDSSNTGAKALTLGLSTGSTSQNR     175
QY      314 SIINPG-TAMASHKDDEDKFFPMSCGMIFGESASANTALDNV-----MITDEEEK     365
          ::||:|:|:|
Db      176 ISLRPGPVSOPLYHMDTDKVVTGINAISHQTTC--NAEDKEYQQGVGRFPNEKEQLK    232
QY      366 ATPVATEREGTVAVNQSSSTDPAIGDVHAMGALPGMVWDRODLYLOGPIWAKIPIHTDG    425
          :.|:|:~|:|:|
Db      233 QLQGLNMHTY----FNKGTOOYTDOIE-RPLWGSWNRRRALHYESQLWSKI-PNLDD    285
QY      426 HFHP-S-FLMGFGLKNRPOLLIKTPVPANPAEFATKFAFITQYSTGOVSVEIEWE     484
          |::|||
Db      286 SFKTGFALGCGWHQHQBEPQI-----FKYYHKVGQLEVLMQWE               323
QY      485 LOKENSK-----RWNP 495
          -|-|--|
Db      324 LLPFMPWELQLHLNWGP 341

RESULT 12
5223424-13
; Patent No. 5223424
; APPLICANT: COCHRAN, MARK; CHIANG, CHRISTINA H.; MACDONALD, RICHARD D.
; TITLE OF INVENTION: ATTENUATED HERPESVIRUSES AND HERPESVIRUSES WHICH INCLUDE FOREIGN DNA ENCODING AN AMINO ACID SEQUENCE
; NUMBER OF SEQUENCES: 16
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/225,032
; FILING DATE: 27-JUL-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 78,519
; FILING DATE: 27-JUL-1987
; APPLICATION NUMBER: 933,107
; FILING DATE: 20-NOV-1986
; APPLICATION NUMBER: 902,887
; FILING DATE: 02-SEP-1986
; APPLICATION NUMBER: 887,140
; FILING DATE: 17-JUL-1986
; APPLICATION NUMBER: 823,102
; FILING DATE: 27-JAN-1986
; APPLICATION NUMBER: 773,430
; FILING DATE: 06-SEP-1985
; SEQ ID NO:13:
; LENGTH: 579
5223424-13

Query Match           7.2%; Score 209.5; DB 6; Length 579; Best Local Similarity 22.7%; Pred. No. 1.1e-10; Matches 138; Conservative 80; Mismatches 237; Indels 153; Gaps 32;

QY      3 SGGAPMADNNEGADVGNASGNMH--CDSTWLGD---RVITTSTRTWALPTYNNHLYKQ 57
          |||||         |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      24 SGGGGG-GGGRGAGVGAVTGTSFNNQTEFOYLGEGLVKITAHASRLIHLMPEHETTYKR 82
          |||||         |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

QY      58 IS--SASTGAS---NDNHYFGYSTPWGYFDENRFCHFSPRDWQRLINNNGFREPKLIN 111
          |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      83 IHVLNSESGVAGQMVDDAHQTOMVTPMWSLIDRNAMGWPFNPADWOILSNMTEINVSEE 142
          |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

QY      112 EKLFIHQVKEVTT--TNDEVTTIANNLSTVOVESDSEYQLPYVLGSAHGCLPPPADV 168
          ::||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
```



Db 143 OEIENVVLKTTTESATSPSPSKIYNNDLTASLWALDTNNTLPYTPAPRSETLGFYPLP 202

QY 169 FMIPQGYL-----TLNNGSQA VGRS-----SFYCLE-YFP SQMLRTGNF 208

Db 203 TKPTQYRYLSCIRNLNPTTYTGOSQITDSIQTLHSDIMFYTIENAVPIHLRTGDEF 262

QY 209 -TPSYTFEEVFPFSSSYAHSQSLDRLMPLIDQYLYLNRTQNSGSAQNKDLLFSRSPA 267

Db 263 STGIYHFDTKPL--KLTHSWQTR-----SLGLPP 290

QY 268 GMSVQP-----KNWLPGPCYRQ--QVSKTKTD-----NNSNFTWTGAS 305

Db 291 KVLTEPTTEGDHPTLPGANTRKGYHQITINNSYTEATALRPAQGVNTPYMNFDYSNGG 350

QY 306 KYNLNGRESINP--GTAMASHKDEDEKFFPMGVMIFGKESAGASNTALDNVITDEEE 363

Db 351 PF-----LTPIVPTADTQYDDE--PNGAIRFTMGYQHGLTTS-----SOLE 392

QY 364 IKATNP-----VATERFGTVA-VNFOSSS-----TDPATG--DVHANGALP-----GM 403

Db 393 RYTFNPQSKCGRAPKQGFNQAPLNLNTNNGTLLPSDPIGKSNKHFNMNTLNTYGPLTA 452

QY 404 VWQDRDYYLQPIWAKIPHTD--GHFHPSPLMGFGLK--NPPQILIKTVPANPAEF 460

Db 453 LNTAPVFPNGQIMDKELDTDLKPLRH--VAPFVCKNNPQGLFVKLAP--NLTDGF 506

QY 461 SA--TKFASFTQYSTGQVSVEIEMELQKENSKRWNPEVQYTSNYAKSANVDFTVDNGL 518

Db 507 NADSPQOPRIITD-SNFWWKGTLTFTAKMRSSNMNPIQOHTT-----TAENIRK 555

QY 519 YTEPRPIG 526

Db 556 YI-PTNIG 562

## RESULT 13

US-09-022-949-2

; Sequence 2, Application US/09022949

; Patent No. 6187759

; GENERAL INFORMATION:

; APPLICANT: Tarpey, Ian

; APPLICANT: Greenwood, Neil

; TITLE OF INVENTION: Canine parvovirus DNA vaccination

; NUMBER OF SEQUENCES: 2

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Akzo No. 6187759el Patent Dept.

; STREET: 1300 Piccard Drive, Suite 206

; CITY: Rockville

; STATE: Maryland

; COUNTRY: US

; ZIP: 20850

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/022, 949

; FILING DATE: 11-FEB-1998

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Gormley, Mary E.

; REGISTRATION NUMBER: 34,409

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 301-948-7400

; TELEFAX: 301-948-9751

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 584 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-09-022-949-2

## Query Match

Best Local Similarity 20.4%;

Matches 130; Conservative 91; Mismatches 223; Indels 192; Gaps 33;

QY 5 GGAPMADNNEGADGVNAGSNWHDSTWLGDRIITST-----RTWALPTYNN- 52

Db 10 GGQP-AVNERATGSGNGSGGGGGS--GGVGISTGTENNQTEFKLENGWEITANGS 66

QY 53 ---HL-----YKQI-----SSASTGASNDNHFGYSTPWCYFEDNRFHCHFSPRDW 95

Db 67 RLVLHNMPESENRYRNVVNNLDKTAVNGMALDDTHAQIVTPWSLYDANAMGWENPGDW 126

QY 96 QRLNNNWGFRPKRLNFKLFIQYKEVT--TNDGVTIANNLSTVQVFSDEYQLPYV 152

Db 127 QLIYNTMSELHVSFEQEIFNVVLKTVSESATQPTKYVNNDLTASLWALDSNNTMPFT 186

QY 153 LGSAGQGL-----PPFPA-----DVFMIPQGYLTLNNGSQA VGRS----- 189

Db 187 PAAMRSETLGFYPWKPTIPTPWRYFQWDRTLIPSH-----TGTSGTPTNIYHGTDPDD 240

QY 190 -SFYCLE-YFP SQMLRTGNF--TPSYTFEEVFPFSSSYAHSQSLDRLMPLIDQYLYLNR 246

Db 241 VQFYTTIENSVPVHLRLRTGDEFATGTFEFDCKP--CRLTHWQTRALG--LPFLNSLPQ 296

QY 247 TQNGS-----GSAQNK-----DLFSRGSPPAGMSVQPKNWLPGPCYRQORV 287

Db 297 SEGTFNFGYIGVQQDKRGGVQMGNTNYITEATIMRPAEYGS-----APYYSFEAS 348

QY 288 SK-----TKTDNNSNFTWGASKY--NLNGRESIINPGT---AMASHKD 327

Db 349 TQGFPRFTPIAAGRGAGQTDENQAA---DGDPRYAFGRQHGQKTTTGTETPERFTYIAHQD 405

QY 328 DEDKFFPMGVMIFGKESAGASNTALD-NVMITDEEIKATNPVATERFGTVA VNFQSSS 386

Db 406 T-----GRYPEGDWTONINENLPTNDNVLLPTDPIG---GKTGINY--TN 446

QY 387 TDPATGDVHANGALPGMWQDRDYYLQPIWAKIPHTDGHFHPSPLMGFGLK----- 439

Db 447 IFNTYGPLTALNNVP-----PYVNGQIMDKEDFTD-----LKPRHLVNA 486

QY 440 -----NPPQILIKTVPAN---PPAEFSATKFASTITQYSTGQVSVEIEMELQKENS 490

Db 487 PFVCCNNCGQLFVKVAPNLINEXYDPDASANMSR---IVTYSDFWKKGLVFKAKLRAS 542

QY 491 KRWNPVQYTSNYAKSANVDFTVDNNGLYTEPRPIG 526

Db 543 HTWNPFIQMSIN-----IDNQFNVYV-PSNIG 567

## RESULT 14

US-08-856-841-17

; Sequence 17, Application US/08856841

; Patent No. 6274307

; GENERAL INFORMATION:

; APPLICANT: ERWIN SOUTSCHEK

; APPLICANT: MANFRED MOTZ

; TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PEPTIDES

; TITLE OF INVENTION: OR POLYPEPTIDES FROM THE PARVOVIRUS B19

; NUMBER OF SEQUENCES: 28

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: BROOKS HAIDT HAFNER &amp; DELAHUNTY

; STREET: 99 PARK AVENUE

; CITY: NEW YORK

; STATE: NY

; COUNTRY: USA

; ZIP: 10016

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5" FLOPPY DISC

; COMPUTER: AT&amp;T - IBM COMPATIBLE

; OPERATING SYSTEM: MS-DOS Version 6.2

; SOFTWARE: ASCII

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER:  US/08/856,841
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:  US/08/214,658
FILING DATE:  16-MARCH-1994
APPLICATION NUMBER:  US 07/917,096
FILING DATE:  4-AUGUST-1992
APPLICATION NUMBER:  PCT/DE91/00106
FILING DATE:  8-FEBRUARY-1991
APPLICATION NUMBER:  DE40038262
FILING DATE:  8-FEBRUARY-1990
ATTORNEY/AGENT INFORMATION:
NAME:  ROBINSON, WILLIAM R.
REGISTRATION NUMBER:  27,224
REFERENCE/DOCKET NUMBER:  LKR-9222-A
TELECOMMUNICATION INFORMATION:
TELEPHONE:  (212) 697-3355
TELEFAX:  (212) 557-5635
TELEX:  NONE
INFORMATION FOR SEQ ID NO:  17:
SEQUENCE CHARACTERISTICS:
LENGTH:  387
TYPE:  AMINO ACID
TOPOLOGY:  LINEAR
MOLECULE TYPE:
DESCRIPTION:  PEPTIDE
HYPOTHETICAL:  N/A
ANTI-SENSE:  N/A
FRAGMENT TYPE:  INTERNAL
ORIGINAL SOURCE:  SERUM FROM PATIENT WITH ACUTE
ORIGINAL SOURCE:  INFECTION (ERYTHEMA INFECTIONOSUM)
IMMEDIATE SOURCE:  GENETICALLY ENGINEERED PEPTIDE
POSITION IN GENOME:  N/A
FEATURE:
NAME/KEY:  N/A
LOCATION:  N/A
IDENTIFICATION METHOD:  amino acid analysis and
IDENTIFICATION METHOD:  mass spectrometry
OTHER INFORMATION:
PUBLICATION INFORMATION:
AUTHORS:  COSSART, Y.E.
AUTHORS:  FIELD, A.M.
AUTHORS:  CANT, B.
AUTHORS:  WIDOWS, D.
TITLE:  PARVOVIRUS-LIKE PARTICLES IN HUMAN SERA
JOURNAL:  LANCET
VOLUME:  I
ISSUE:
PAGES:  72 - 73
DATE:  1975
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO:  17:
PUBLICATION INFORMATION:
AUTHORS:  MANIATIS, T.
AUTHORS:  FRITSCH, E.F.
AUTHORS:  SAMBROOK, J.
TITLE:  MOLECULAR CLONING
JOURNAL:  COLD SPRING HARBOR, NY
VOLUME:
ISSUE:
PAGES:
DATE:  1982
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO:  15:
PUBLICATION INFORMATION:
AUTHORS:  SMITH, D.B.
AUTHORS:  JOHNSON, K.S.

```

; TITLE: SINGLE STEP PURIFICATION OF POLYPEPTIDES  
 ; TITLE: EXPRESSED IN ESCHERICHIA COLI AS FUSIONS WITH  
 ; TITLE: GLUTATHIONE S. TRANSFERASE  
 ; JOURNAL: GENE  
 ; VOLUME:  
 ; ISSUE: 67  
 ; PAGES: 31 - 40  
 ; DATE: 1988  
 ; DOCUMENT NUMBER:  
 ; FILING DATE:  
 ; PUBLICATION DATE:  
 ; RELEVANT RESIDUES IN SEQ ID NO: 17:  
 ;  
 US-08-856-841-17

```

Query Match          5.9%; Score 171; DB 4; Length 387;
Best Local Similarity 33.1%; Pred. No. 2e-07;
Matches 43; Conservative 17; Mismatches 58; Indels 12; Gaps 2

QY      10 ADNNEGADVGNGAS--GNWHCDSTWLGDRVITTTSTRTWALPTYNHLLYKQISSASTGASN 67
      ||: ||| ||: | :|: | :|: | :|: | :|: |
Db      242 AEASTGAGGGGNSVKSWMSEGAFTSANSVCTCFSRQFLIPYDPEHHYKVFSPAASSCHN 301

QY      68 D-----NHYFGYSTPWGYFDENRFHCHFSPRDMQRLINNNGWGRPKRLNFKLFNI 117
      : ||||| | ||| : ||| : ||| : ||| : ||| :
Db      302 ASGKEAKVCTISPIMGYSIPWRYYLDENALNLFSPLEFQHLIENYGSIAPDALTVTISEI 361

QY      118 QVKEVTNDG 127
      ||: ||| |
Db      362 AVKDVTDKTG 371

```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
US-08-469-202-27

Query Match 3.8%; Score 111; DB 1; Length 655;  
Best Local Similarity 18.6%; Pred. No. 0.14;  
Matches 119; Conservative 77; Mismatches 229; Indels 214; Gaps 32;

QY 8 PMADNNEGADGVGNAS--GNWCHDSTWLGDRV-----ITTSTRTWALPTYNNHLYKQISS 60  
DB 104 PPIDNVNNTDAAGNTGYHGYWGRDYFRIDHFGNLDDEKELTSLMSPDYNMKLVLDYAP 163  
QY 61 ASTGASNDNH-----FGYSTPWGYFDENRFCHFSPPRDWQRLINNMG 104  
DB 164 NNSNANDENEFGLYRDGVFITDYPTDVAANTGW-----YHNGGVYTNW-----NDF 210  
QY 105 FRPKRLNFKLFNIQVKEVTTNDGVTIANNLJSTVQVFSDEYQLPYVLGSAH--QGCLP 162  
DB 211 FOVK--NHNLFNLS---DLNQSNTDVQYLLDGSKEWIDAGVDAIRIDAIKHMDSFIQ 264  
QY 163 PFPADVFMIPQYGYLTLNNGSOAVGRSSFYCL-EYFPSQMLRT-----GN----- 206  
DB 265 KWTSDI-----YDY-----SKSIGREGFFPFGHEWFGASANTTTGVDGNAIDYANTS 313  
QY 207 --NFTFSYTFEEVPHSSYAHQSOLDRLMPLIDQYLYL----- 244  
DB 314 LLDGFRDITLERVLVGRS--GNTMKTLSYLIKQTVFTSDWQVFMNDHDMARIGTA 370  
QY 245 ---NRT-----QNSGSAQNKDLLFSR---GSPAGMSVQPKNWLPGPCYRQQRVSKTKTD 293  
DB 371 LRSNATTFGPNNETGGSQSEAFQAKRIDLGIVATMTVR--GIPAIYGTGHEVYANFTS 427  
QY 294 NNSNFTWTGASKYNLNGRESIINPGTAMASHKDEDEKFFPM----- 335  
DB 428 N---SFGQVGSDPYN---REKM--PGF-----DTESEAFSIKTLGLRKSSPAIONGT 473  
QY 336 -----SGVMIFGKESA-----GASNTA-LDNVMTD-----EEEIKATNP 369  
DB 474 YTELMVNDLILVERRSNDIVIVALNRGEANTINVKNIHAVENGVPSPSIGNNSVSVANK 533  
QY 370 VAT-ERFGTVAVNQSSSTDPAIGDVHAM--GALPGMWQDRDVIYLOGPIWAKIPHTDGH 426  
DB 534 QATLTMQNEAVVIRSQSDDAENPTVQSIINPACNNGYTTISGQSVYIIGNI----- 583  
QY 427 FHPSPLMGGFGLKNPPQILIKNTPVPPANPPAERSATKFAFITQYSTGQVSVE----- 480  
DB 584 ---POLGGWDLTK-----AVKISP-----TQYPOWSASLELPSDLN 616  
QY 481 IEMELQKENSKRWNPEVQYTSNYAKSANVDFVNNGLY 519  
DB 617 VEMKCVKRNETNPANTVEMQSGANNQFNSNDTQTNGSF 655

Search completed: July 17, 2003, 18:36:39  
Job time : 14.1429 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 2003, 18:34:40 ; Search time 25.7143 Seconds  
(without alignments)  
2466.245 Million cell updates/sec

Title: US-09-807-802A-17  
Perfect score: 2906  
Sequence: 1 MASGGAPMADNNEGADVG.....NNGLYTEPRPIGTRYLTRPL 534

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 451899 segs, 118759770 residues

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:\*

1: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep1:\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep2:\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep3:\*  
13: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2481.5	85.4	533	15	US-10-038-972A-15 Sequence 15, Appl
2	2481.5	85.4	598	15	US-10-038-972A-14 Sequence 14, Appl
3	2481.5	85.4	735	15	US-10-038-972A-13 Sequence 13, Appl
4	2481.5	85.4	735	15	US-10-293-478-1 Sequence 1, Appl1
5	491.5	16.9	756	15	US-10-205-942-4 Sequence 4, Appl1
6	108	3.7	5795	10	US-09-815-242-12610 Sequence 12610, A
7	104.5	3.6	717	15	US-10-046-583A-1 Sequence 1, Appl1
8	104	3.6	2344	10	US-09-815-242-12713 Sequence 12713, A
9	103.5	3.6	1016	10	US-09-815-242-5845 Sequence 5845, Ap
10	103.5	3.6	1265	15	US-10-198-070-69 Sequence 69, Appl
11	102	3.5	3712	15	US-10-108-605-103 Sequence 103, App
12	101.5	3.5	439	9	US-09-836-353A-145 Sequence 145, App
13	101.5	3.5	439	12	US-09-984-130-145 Sequence 145, App
14	101	3.5	1411	15	US-10-080-505-17 Sequence 17, Appl
15	100.5	3.5	1714	15	US-10-098-916A-11 Sequence 11, Appl
16	99	3.4	824	15	US-10-046-583A-3 Sequence 3, Appl1

17	98.5	3.4	1000	15	US-10-055-412B-30	Sequence 30, Appl
18	97.5	3.3	685	10	US-09-745-763-11	Sequence 11, Appl
19	97	3.3	659	15	US-10-046-583A-2	Sequence 2, Appl1
20	97	3.3	1844	15	US-10-242-056-53	Sequence 53, Appl
21	97	3.3	2504	15	US-09-817-514A-8	Sequence 8, Appl1
22	97	3.3	2504	15	US-10-242-056-12	Sequence 12, Appl
23	96	3.3	1394	12	US-09-839-996-2	Sequence 2, Appl1
24	96	3.3	1394	15	US-10-080-505-2	Sequence 2, App
25	95.5	3.3	1286	15	US-10-153-668-212	Sequence 212, App
26	95.5	3.3	1354	15	US-10-153-668-470	Sequence 470, App
27	95	3.3	433	11	US-09-859-888-4	Sequence 4, Appl1
28	95	3.3	478	11	US-09-859-888-2	Sequence 2, Appl1
29	95	3.3	1545	12	US-09-839-996-4	Sequence 4, Appl1
30	95	3.3	1545	15	US-10-080-505-4	Sequence 4, Appl1
31	95	3.3	2150	15	US-10-135-322-17	Sequence 17, Appl
32	94.5	3.3	462	15	US-10-127-032-101	Sequence 101, App
33	94.5	3.3	831	10	US-09-732-665-7	Sequence 7, Appl1
34	94	3.2	1395	15	US-10-080-505-7	Sequence 7, Appl1
35	94	3.2	4590	15	US-10-160-758-13	Sequence 13, Appl
36	94	3.2	4590	15	US-10-160-758-14	Sequence 14, Appl
37	94	3.2	4590	15	US-10-060-036-157	Sequence 157, App
38	93.5	3.2	1156	10	US-09-826-660-15	Sequence 15, Appl
39	93.5	3.2	1178	10	US-09-851-194-2	Sequence 2, Appl1
40	93.5	3.2	1770	10	US-09-841-132-444	Sequence 444, App
41	93.5	3.2	1848	12	US-09-839-996-6	Sequence 6, Appl1
42	93.5	3.2	1848	15	US-10-080-505-6	Sequence 6, Appl1
43	93	3.2	1404	10	US-09-811-045A-1	Sequence 1, Appl1
44	92.5	3.2	716	12	US-09-349-058-38	Sequence 38, Appl
45	92.5	3.2	716	15	US-10-040-430-38	Sequence 38, Appl

ALIGNMENTS

RESULT 1  
US-10-038-972A-15  
; Sequence 15, Application US/10038972A  
; Publication No. US20020192823A1  
; GENERAL INFORMATION:  
; APPLICANT: J. Bartlett  
; TITLE OF INVENTION: AAV VECTORS AND METHODS  
; FILE REFERENCE: 28335/36996US  
; CURRENT APPLICATION NUMBER: US/10/038, 972A  
; CURRENT FILING DATE: 2002-01-04  
; PRIOR APPLICATION NUMBER: US 60/260, 124  
; PRIOR FILING DATE: 2001-01-05  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 15  
; LENGTH: 533  
; TYPE: PRT  
; ORGANISM: adeno-associated virus 2 VP3 capsid protien  
US-10-038-972A-15

QY	1	MASGGAPMADNNEGADVG	85.4%;	Score 2481.5;	DB 15;	Length 533;
		Best Local Similarity	83.3%;	Pred. No. 7.5e-225;		
		Matches 445;	Conservative 37;	Mismatches 51;	Indels 1;	Gaps 1;
QY	1	MASGGAPMADNNEGADVG	85.4%;	Score 2481.5;	DB 15;	Length 533;
		Best Local Similarity	83.3%;	Pred. No. 7.5e-225;		
		Matches 445;	Conservative 37;	Mismatches 51;	Indels 1;	Gaps 1;
QY	1	MATSGAPMADNNEGADVG	85.4%;	Score 2481.5;	DB 15;	Length 533;
		Best Local Similarity	83.3%;	Pred. No. 7.5e-225;		
		Matches 445;	Conservative 37;	Mismatches 51;	Indels 1;	Gaps 1;
QY	61	ASTGASNDNHFGYSTP	85.4%;	Score 2481.5;	DB 15;	Length 533;
		Best Local Similarity	83.3%;	Pred. No. 7.5e-225;		
		Matches 445;	Conservative 37;	Mismatches 51;	Indels 1;	Gaps 1;
QY	61	QSGASNDNHFGYSTP	85.4%;	Score 2481.5;	DB 15;	Length 533;
		Best Local Similarity	83.3%;	Pred. No. 7.5e-225;		
		Matches 445;	Conservative 37;	Mismatches 51;	Indels 1;	Gaps 1;
QY	121	EVTNDGVTITIANLTS	85.4%;	Score 2481.5;	DB 15;	Length 533;
		Best Local Similarity	83.3%;	Pred. No. 7.5e-225;		
		Matches 445;	Conservative 37;	Mismatches 51;	Indels 1;	Gaps 1;
QY	120	EVTQNDGTTITIANL	85.4%;	Score 2481.5;	DB 15;	Length 533;
		Best Local Similarity	83.3%;	Pred. No. 7.5e-225;		
		Matches 445;	Conservative 37;	Mismatches 51;	Indels 1;	Gaps 1;
QY	181	NGSQAVGRSSFYCLE	85.4%;	Score 2481.5;	DB 15;	Length 533;
		Best Local Similarity	83.3%;	Pred. No. 7.5e-225;		
		Matches 445;	Conservative 37;	Mismatches 51;	Indels 1;	Gaps 1;
QY	180	NGSQAVGRSSFYCLE	85.4%;	Score 2481.5;	DB 15;	Length 533;
		Best Local Similarity	83.3%;	Pred. No. 7.5e-225;		
		Matches 445;	Conservative 37;	Mismatches 51;	Indels 1;	Gaps 1;

OY	241	LYYLNRTONQSGSAQNKDLLFSRGS	PAGMSVOPKNWLPGPCYRQORVS	KTKTDNNNSNFT	300
Db	240	LYLSRTNPSTGTTQSRLQFSQAGASD	IRDQSRNWLPGPCYRQORVSKTSAD	NNNSEYS	299
OY	301	WTGASKYNLNGRESIINPGTAMASHK	DEDEKEFPMSGWIMFGESAGASANTALDN	MITD	360
Db	300	WTGATKYHLNGRDSLVPNPGPAMASHK	DEEKEFPQSGVLIFGQSEKTVDIEKM	ITD	359
OY	361	EETKATNPVATERFGTVAVNFQSS	STDPA TG DVHAMGALPGVMWD R D V Y L	OGPIWAKI	420
Db	360	EELRTNPVATEQYGSVSTNLQRGNRQA	TADVNTQGVLPGVMWQDRDYVL	OGPIWAKI	419
OY	421	PHTDGHFHPSPLMGGFGLKNPPQIL	IKNTPYPANPRAEFSA TKFASFITQYSTG	QVSVE	480
Db	420	PHTDGHFHPSPLMGGFGLKHPPQIL	IKNTPEPANPSTTFSAKAFSFI T QYSTG	QVSVE	479
OY	481	IEMELOKENS KRW N P E VOYT S N Y A K S A N V D F T V D N N G L Y T E P R P I G T R Y L T R P L		534	
Db	480	IEMELOKENS KRW N P E IOYT S N Y N K S V A N D F T V D N G V Y S E P R P I G T R Y L T R N L		533	

```

RESULT 2
US-10-038-972A-14
; Sequence 14, Application US/10038972A
; Publication No. US20020192823A1
; GENERAL INFORMATION:
; APPLICANT: J. Bartlett
; TITLE OF INVENTION: AAV VECTORS AND METHODS
; FILE REFERENCE: 28335/36996US
; CURRENT APPLICATION NUMBER: US/10/038,972A
; CURRENT FILING DATE: 2002-01-04
; PRIOR APPLICATION NUMBER: US 60/260,124
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
; LENGTH: 598
; TYPE: PRT
; ORGANISM: adeno-associated virus 2 VP2 capsid protien
US-10-038-972A-14

```

Query Match	85.4%;	Score 2481.5;	DB 15;	Length 598;
Best Local Similarity	83.3%;	Pred. No. 8.9e-225;		
Matches 445;	Conservative 37;	Mismatches 51;	Indels 1;	Gaps 1;

QY	1	MASGGAPMADNNEGADGVGNASGNWCHDSITWLGDRIVITTTSTRTWALPTYNNHLYKQISS	60
Db	66	MATGSGAPMADNNEGADGVGNSSGNWCHDSITWNGDRVITTTSTRTWALPTYNNHLYKQISS	125
QY	61	ASTGASNDNHFFGYSTPMGWYFDENRFFHCHFSERDWRLLINNMGFRPKRLNFKLENIQVK	120
Db	126	QS-GASNDNHFFGYSTPMGWYFDENRFFHCHFSERDWRLLINNMGFRPKRLNFKLENIQVK	184
QY	121	EYTTNDGVTTIANNLTSTVQVFSDSERYQLPYVLGSAHQGLPPFPADVFMIPOYGYLTIN	180
Db	185	EYTONDGTITIANNLSTVQVFTDSEYQLPYVLGSAHQGLPPFPADVFMIPOYGYLTIN	244
QY	181	NGSOAVGRSSFYCLBYFPSQMLRTGNFTFESYTFEEVPFHSSYAHQSGLDRLMNPILIDY	240
Db	245	NGSOAVGRSSFYCLBYFPSQMLRTGNFTFESYTFEDVPFHSSYAHQSGLDRLMNPILIDY	304
QY	241	LYLINRTONOGSSAQNKDLLFSRGSFAGMSVQPKWMLPGPCYRQORVSKTKTDNNNSNFT	300
Db	305	LYLISRTNTPSGTTTQSRLOFQSQAGASDIRDOSRNWMLPGPCYRQORVSKTSADNNNSEYS	364
QY	301	WTGASKYNLNGRESIINPGTAMASHKODEDEKFFPMGSMIFGKESAGASNTALDNVMITD	360
Db	365	WTGATKYHLNGRDSLVPNGPAMASHKODEEKFFPQSGVLIFGQOGSEKTNVDIEKVMITD	424
QY	361	EEELIKATNPVATERFGTVAVNFOSSSTDPAIGDVHAMGALPGMWQDRDVILOGPIMAKI	420
Db	425	EEELIRTNPEVATEOYGSVSSTNLORGNGROATADVNTQOVLPGMWQDRDVILOGPIMAKI	484

QY	421	PHTDGHFHPSPLMGGFGLKNPPQILLKNTVPANPPAEFSATKFASTIQYSTGOVSVE	480
Db	485	PHTDGHFHPSPLMGGFGLKNPPQILLKNTVPANPSTTFSAKFASTIQYSTGOVSVE	544
QY	481	IEMELQENSKRMNPEVOYTSNYSAKSANDVFTVDNNGLYTEPRPIGTRYLTRPL	534
Db	545	IEMELQENSKRMNPEIQYTSNYSKSNVNDFTVDINGVYSEPRPIGTRYLTRNL	598

```

RESULT 3
US-10-038-972A-13
; Sequence 13, Application US/10038972A
; Publication No. US20020192823A1
; GENERAL INFORMATION:
; APPLICANT: J. Bartlett
; TITLE OF INVENTION: AAV VECTORS AND METHODS
; FILE REFERENCE: 28335/36996US
; CURRENT APPLICATION NUMBER: US/10/038,972A
; CURRENT FILING DATE: 2002-01-04
; PRIOR APPLICATION NUMBER: US 60/260,124
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1.
; SEQ ID NO 13
; LENGTH: 735
; TYPE: PRT
; ORGANISM: adeno-associated virus 2 VP1 capsid protien
US-10-038-972A-13

```

Query Match	85.4%	Score 2481.5;	DB 15;	Length 735;
Best Local Similarity	83.3%	Pred. No. 1.2e-224;		
Matches 445;	Conservative 37;	Mismatches 51;	Indels 1;	Gaps 1;

QY	1	MASGGGAPMADNNEGADGVGNASGNMHCDSWLGDRIVTTSTRTWALPTYNHLYKQISS	60
Db	203	MATGSGAPMADNNEGADGVGNSSGNMHCDSWTMGDRVITTTSTRTWALPTYNHLYKQISS	262
QY	61	ASTGASNDNHYFGYSTPWGYFDENRFCHFSPRDWORLINNNGFRPKRLNEKLENIQVK	120
Db	263	OS-GASNDNHYFGYSTPWGYFDENRFCHFSPRDWORLINNNGFRPKRLNEKLENIQVK	321
QY	121	EVTNDGVTTIANNLSTQVQSEDSYQLPYVLSAHQGLPFPADVFMIPQYGYLTN	180
Db	322	EVTQNDGTTIANNLSTQVQFTDSEYQLPYVLSAHQGLPFPADVFMIPQYGYLTN	381
QY	181	NGSQAVGRSSFYCLEYFPSQMLRTGNNTFSYTFEEVPFHSSYAHSQSLDRMLNPLIDQY	240
Db	382	NGSQAVGRSSFYCLEYFPSQMLRTGNNTFSYTFEDVPFHSSYAHSQSLDRMLNPLIDQY	441
QY	241	LYTLNRTQNSGSAQNKDLFRGSPAGMSVQPKNMLPGPCYRQORVSKTKTDNNNSNFT	300
Db	442	LYTLRTNTPSGTTTQSRLOFSQAGASDIRDQSRNMLPGPCYRQORVSKTSADNNNSEYS	501
QY	301	WTGASKYNLNGRESIINPGTAMASHKODEDEKFFPMGVMIFGKESAGASNTALDNVMTTD	360
Db	502	WTGATKYHLNGRDSLVPNPGPAMASHKODEKEFFPQSGVLI FGKQSEKTVNIEKVMITD	561
QY	361	EEBIKATNPVATERFGTVAVNFQSSSTDPAATGDVHAMGALPGMWQDRDVILOCFIWAKI	420
Db	562	EEHIRTNPVATEQYGSVSTNLQGRNROAATADVNTQGVLPGMWQDRDVILOCFIWAKI	621
QY	421	PHTDGHFHPSPMLMGFGGLKNPPOILIKNTPVPANPPAEFSATKFASTFOYSTGOVSVE	480
Db	622	PHTDGHFHPSPMLMGFGGLKHPPOILIKNTPVPANPSTTFSAAKFASTFOYSTGOVSVE	681
QY	481	IEMELOKENSKRWNPEVOYTSNYAKSANDFTVDNNGLYTEPRPIGTRYLTRPL	534
Db	682	IEMELOKENSKRWNPEIQTYSNYSNVNDFTVDTNGVYSEPRPIGTRYLTRNL	735

RESULT 4  
US-10-293-478-1

```
; Sequence 1, Application US/10293478
; Publication No. US20030078411A1
; GENERAL INFORMATION:
; APPLICANT: PATEL, SATIL D.
; APPLICANT: MCARTHUR, JAMES G.
; TITLE OF INVENTION: ADENO-ASSOCIATED VIRUS CAPSID IMMUNOLOGIC DETERMINANTS
; FILE REFERENCE: 39672
; CURRENT APPLICATION NUMBER: US/10/293,478
; CURRENT FILING DATE: 2002-11-14
; PRIOR APPLICATION NUMBER: US/09/321,589
; PRIOR FILING DATE: 1999-05-28
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 735
; TYPE: PRT
; ORGANISM: Adeno-associated virus
US-10-293-478-1

Query Match      85.4%; Score 2481.5; DB 15; Length 735;
Best Local Similarity 83.3%; Pred. No. 1.2e-224;
Matches 445; Conservative 37; Mismatches 51; Indels 1; Gaps 1;

QY      1 MASGGGAPMADNNEGADGVGNASGNWHDSTWLGDRVITTTSTRTWALPTYNHLYKQISS 60
DB      203 MATGSGAPMADNNEGADGVGNSSGNWHDSTWMDRVITTTSTRTWALPTYNHLYKQISS 262
QY      61 ASTGASNDNHFGYSTPWGTFDFNRFHCHFSPRDWORLNNWGFPRKRLNFKLFNIQVK 120
DB      263 QS-GASNDNHFGYSTPWGTFDFNRFHCHFSPRDWORLNNWGFPRKRLNFKLFNIQVK 321
QY      121 EYTTNDGVTTIANNLSTVQVFSDEXQLPYVLGSAHQCLPPPADVEMIPQYGYLTIN 180
DB      322 EYTONDGVTTIANNLSTVQVFTDSEYQLPYVLGSAHQCLPPPADVEMIPQYGYLTIN 381
QY      181 NGSQAVGRSSFYCLEYFSPQMLRTGNNTFSYTFEEVPHSSYAHSQSLDRLMPLIDQY 240
DB      382 NGSQAVGRSSFYCLEYFSPQMLRTGNNTFSYTFEEVPHSSYAHSQSLDRLMPLIDQY 441
QY      241 LYYLNRTQNSGSAQNKDLFSRGSPPAGMSVQPKNWLPGPCYRQORVSKTKTDNNNSNFT 300
DB      442 LYYLSRTNTPSGTTTQSRLOFSQAGASDIRQSRNWLPGPCYRQORVSKTSADNNSEYS 501
QY      301 WTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMGVMIFGKESAGASNTALDNVMTD 360
DB      502 WTGATKYNLNGRDSLVPNPGPAMASHKDEDEKFFQSGVLI FGKQSEKTNVDIEKVMITD 561
QY      361 EEEIKATNPVATERFGTVAVNFOSSSTDPATGDVHAMGALPGMWQDRDVLVLOGPIWAKI 420
DB      562 EEEIRTPNPVATEQYGSVSTNLQGRNQATADVNTQGVLPGMVQDRDVLVLOGPIWAKI 621
QY      421 PHTDGHFHPSPLMGFGFKNPPQILIKNTVPVNPANPPAEFSATKFAFITQYSTGVSV 480
DB      622 PHTDGHFHPSPLMGFGFKNPPQILIKNTVPVNPANPPSTTFSAKFAFITQYSTGVSV 681
QY      481 IEWELOKENSKRWNPEVQYTSNYAKSANDFTVDNNGLYTEPRPIGTRYLTRPL 534
DB      682 IEWELOKENSKRWNPEIQYTSNYKNSVNDFTVDNNGVYSEPRPIGTRYLTRNL 735

RESULT 5
US-10-205-942-4
; Sequence 4, Application US/10205942
; Publication No. US20030053990A1
; GENERAL INFORMATION:
; APPLICANT: University of No. US20030053990A1ch Carolina-Chapel Hill
; APPLICANT: Rabinowitz, Joseph E.
; APPLICANT: Samuleki, Richard J.
; APPLICANT: Xiao, Weidong
; TITLE OF INVENTION: VIRUS VECTORS AND METHOD OF MAKING AND ADMINISTERING THE SAME
; FILE REFERENCE: 5470-186
; CURRENT APPLICATION NUMBER: US/10/205,942
; PRIOR FILING DATE: 2002-07-26
```

```
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 756
; TYPE: PRT
; ORGANISM: Adeno-associated virus
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(2271)
; OTHER INFORMATION: B19/AAV chimeric capsid coding sequence
US-10-205-942-4

Query Match      16.9%; Score 491.5; DB 15; Length 756;
Best Local Similarity 27.6%; Pred. No. 3e-37;
Matches 144; Conservative 83; Mismatches 239; Indels 55; Gaps 16;

QY      10 ADNNEGADGVGNAS--GNWHDSTWLGDRVITTTSTRTWALPTYNHLYKQISSASTGASN 67
DB      209 AEAATGAGGGGGSNSVKMSWSEGATFSANSVTCTFSRQFLIPYDEPHYKVFSPASSCHN 268
QY      68 D-----NHFGYSTPWGTFDFNRFHCHFSPRDWORLNNWGFPRKRLNFKLFNI 117
DB      269 ASGKEAKVCTISPIMGYSTPWRYLDENALNLFSSPLEFOHLIENYGSIAPDALTITSEI 328
QY      118 QVKEVT--TNDGVTTIANNLSTVQVFSDEXQLPYVLGSAHQCLPPPADVEMIPQY 175
DB      329 AVKDVTDKTTGGV-QVTDSTGRCLMLVDHEKYRPPYVLGGODTLAPBLPIWVFPQYA 387
QY      176 YLTNN-GSQAVG-----RSSFYCLEYFSPQMLRTGNNTFSYTFEEVPHSSYAH 225
DB      388 YLTVDVNTQGISGDSKRLASBSAFLVLEHSSFQLTGTTATWTSYKFPVPPEMLEGC 447
QY      226 SQSLDRLMPLIDQYLYLNRTQNSGSAQNKDLFSRGSPPAGMSVQPKNWLPGPCYRQ 285
DB      448 SQHFYEMYNPL---YGSRLGVPTDLGGDPKFRSL-----THEDHAIQPNFMGPLVNSV 499
QY      286 RVSKTKTDNNNSNFTWTGASKYNLNGRESIINPG-TAMASHKDEDEKFFPMGVMIFGKE 344
DB      500 STKEGDSNTGAKALTGLSTGSRQNTRLISLRGPVSPQPYHMDTDKYVTGINAISHQGT 559
QY      345 SAGASNTALDNV-----MITDEEIKATNPVATERFGTVAVNFOSSSTDPATGDVHAM 397
DB      560 TYG---NAEDKEYQGVGRFPNEKEQLKQLOGLNHTY-----FPNKGTOQYTDQIE-R 609
QY      398 GALPGMWQDRDVLVLOGPIWAKIPHTDGHFHPSP-PLMGFGFKNPPQILIKNTVPVNP 456
DB      610 PLNVGSVWNRRAALHYESQLSMKIPNLDDSFKTQFALGCGWGLHQPPOIFLK--ILPQSG 667
QY      457 PAESATKFAFITQYSTGVSVIEIWELOKENSKRWNPE 496
DB      668 PIGGIKSMGITTLVQYAVGIMVTMTFKLGPRKATGRWNPQ 708

RESULT 6
US-09-815-242-12610
; Sequence 12610, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
```



;; PRIOR APPLICATION NUMBER: 60/206,848  
;; PRIOR FILING DATE: 2000-05-23  
;; PRIOR APPLICATION NUMBER: 60/207,727  
;; PRIOR FILING DATE: 2000-05-26  
;; PRIOR APPLICATION NUMBER: 60/242,578  
;; PRIOR FILING DATE: 2000-10-23  
;; PRIOR APPLICATION NUMBER: 60/253,625  
;; PRIOR FILING DATE: 2000-11-27  
;; PRIOR APPLICATION NUMBER: 60/257,931  
;; PRIOR FILING DATE: 2000-12-22  
;; PRIOR APPLICATION NUMBER: 60/269,308  
;; PRIOR FILING DATE: 2001-02-16  
;; NUMBER OF SEQ ID NOS: 14110  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 12610  
;; LENGTH: 5795  
;; TYPE: PRT  
;; ORGANISM: Staphylococcus aureus  
US-09-815-242-12610

Query Match 3.7%; Score 108; DB 10; Length 5795;

Best Local Similarity 18.7%; Pred. No. 8.6;  
Matches 120; Conservative 85; Mismatches 232; Indels 204; Gaps 34;

QY 5 GGAPMADNNEGADGVGNASGNMHCSTWLGDREVITSTRTWALPTVNNHLYKQISSASTG 64  
DB 1723 GQAPNKDNTRIGEDI-NVTANILID-GETTPIKTAT-----YKVS- 1762  
QY 65 ASNDNHFGYSTFWGYFDENRFCH-----FSPRDWQRLINNMGRPRKLNFKLFNI- 117  
DB 1763 -----SYFKHVFETNRGAVFPVSDVYDAKQYKPVNDSWTQNAQRMNFQFTNSY 1812  
QY 118 ----QVKEVTNDGVTITIANNLSTVQVPSSEYQLPYVLGSA----- 156  
DB 1813 GPSKDVGVGISTRIDIRVTDYDNHQTQIKILAKYKDPDRIDGNSVITYAGLTNQOIKINNV 1872  
QY 157 -HQCCLPPFPADVFMIPQYGLTLNNGSQAVGRSSFYCL-EYFPSQMLRTGNNFTSYTF 214  
DB 1873 LSSSSIKLFKAD--NTP---LTTNTTYGSGNTAVVTVS DALPNGVTKARS---SITM 1922  
QY 215 EEVFPFSSYAHQSOLDRLMNPILIDQYLYLNRTQNSGS-----AQNKDLFSRGS- 265  
DB 1923 NNVTYTTQDEHGRAIDVTRNESVD-----SND SATVTVPOLQATTEGAVFIKGD 1973  
QY 266 -----PAGMSV---QPKNW-----LPGPCY 282  
DB 1974 GFDFGHERFIONPFGATVAMHDNPDFTKNTVGNTHKTA VVTLPSGGGTRNVEVPKVY 2033  
QY 283 RQQRVSKTKTDNNNSNFTWTGASKYNLNGRESI-----INPGT-----AMASHKDEDK 331  
DB 2034 PVANAKAPSRDYKGNLT-----NGTDAINYITFDPNTNTNGITAAWANNRQOPNNQ 2084  
QY 332 FFPMSGVMIFGKESAGASNTALDNVMT---DEEIKATNPVATERFGTVA VNFQSSSTD 388  
DB 2085 ---QAGVQHLNVDTVYPGITAARVPVTVNVYQFEPQTSYTTVG-GTLANGTQ----- 2135  
QY 389 PATGDVHAMGA--LP--GMV--WQDRVYIQQPIWAKI--PHT-----DGHFH 428  
DB 2136 -ASGYAHMQNANGLPTDFTYKMNNAATGTNDANWAAAMNKPAKVNAKYDVIMGHTF 2194  
QY 429 PSPLMGFGGLKN-PPQILIKNTPVPA---NPPEFSATKFAFITQYS-----TGQV 477  
DB 2195 ATSLPAKFVYQVQPAKPTVTETAGAITTEPGANQTVNTHAGVNTYADKLVIKRNQNV 2254  
QY 478 SVEIEWELQKENSKRWNPEVQYTSNYAKSANVDFVDNNGL 518  
DB 2255 ---VTIFTRNNTPWVKE---ASAATVAGTAGT--NNGI 2286

RESULT 7  
US-10-046-583A-1  
; Sequence 1, Application US/10046583A  
; Publication No. US20020168743A1

;; GENERAL INFORMATION:  
;; APPLICANT: Rice, John  
;; APPLICANT: Klotz, Andreas  
;; APPLICANT: Crawford, John  
;; APPLICANT: Lanning, Beth  
;; APPLICANT: Stewart, Sandy  
;; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE  
;; TITLE OF INVENTION: IDENTIFICATION OF MODULATORS OF DEOXYXYLULOSE 5-PHOSPHATE  
;; FILE REFERENCE: 2037 US Divisional  
;; CURRENT APPLICATION NUMBER: US/10/046,583A  
;; PRIOR FILING DATE: 2001-10-26  
;; PRIOR APPLICATION NUMBER: 09/626,589  
;; NUMBER OF SEQ ID NOS: 6  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 1  
;; LENGTH: 717  
;; TYPE: PRT  
;; ORGANISM: Arabidopsis thaliana  
US-10-046-583A-1

Query Match 3.6%; Score 104.5; DB 15; Length 717;

Best Local Similarity 20.4%; Pred. No. 0.76;  
Matches 77; Conservative 54; Mismatches 121; Indels 125; Gaps 18;

QY 122 VTNDGVT---IANNLTSTVQVPSDSEYQLPYVLGSAHQCLPPFPADVFMIPQYGLT 178  
DB 12 IITKGGLSTDCKSTSLSSRSRLVTD---LP-----SPCLKP----- 45  
QY 179 LNNGSQAVGRSSFYCL-----EYF---PSQMLRTGNNFTSYTFEEVFPFSSYAHQS 228  
DB 46 -NNNSHS-NRAKVCASLAKEGEYYSNRPPPTPLDTIN-----YPIHMKNLVKE 93  
QY 229 LDRLMNPILIDQYLYLNRTQNSGSA-----QNKDLFSRGSFAGMSV 271  
DB 94 LKQLSDELRSVDVIFNVSKTGHLGSSLGVELTVALHYIFNTPQDKILW---DVGHQS 148  
QY 272 QPKNWLPGPCYRQQRVSKTKTDNNNSNFTWTGASKYNLNG---RESIINPGTAMASHKD 328  
DB 149 YPHKILTG---RRKMPTRQTNGLSGFTKRGESEHDCFGTGHSTTISAGLMAVGRL 205  
QY 329 EDKFPMSGVMIFGKESAGASNTALDNVMTDEEIKATN-----PVAT-----ERFG 376  
DB 206 KGKNNNVAVIGDAMTAGQAVEAMNAGYLDSDMIVILNDNKQVSLPTATLDGSPSPVG 265  
QY 377 TVAVNFQSSSTDPA-----GDVHAMGA-----LPGMV-----WQDRD 409  
DB 266 ALSSALSRLOSNPALRELREVAKGMTKQIGPMHQLAAKVDYARGMISGTGSSLFEELG 325  
QY 410 VYLGPIWAKIPHTDGH 426  
DB 326 LYITGPV-----DGH 335

RESULT 8  
US-09-815-242-12713  
; Sequence 12713, Application US/09815242  
; Patent No. US20020061569A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Haselbeck, Robert  
;; APPLICANT: Ohlsen, Kari L.  
;; APPLICANT: Zyskind, Judith W.  
;; APPLICANT: Wall, Daniel  
;; APPLICANT: Trawick, John D.  
;; APPLICANT: Carr, Grant J.  
;; APPLICANT: Yamamoto, Robert T.  
;; APPLICANT: Xu, H. Howard  
;; TITLE OF INVENTION: Identification of Essential Genes in  
;; TITLE OF INVENTION: Prokaryotes  
;; FILE REFERENCE: ELITRA.011A  
;; CURRENT APPLICATION NUMBER: US/09/815,242  
;; CURRENT FILING DATE: 2001-03-21



APPLICANT: GEMMELL, JACK  
TITLE OF INVENTION: PEPTIDES EFFECTIVE IN THE TREATMENT OF TUMORS AND OTHER  
TITLE OF INVENTION: CONDITIONS REQUIRING THE REMOVAL OR DESTRUCTION OF  
TITLE OF INVENTION: CELLS  
FILE REFERENCE: 59003.000008  
CURRENT APPLICATION NUMBER: US/10/198,070  
CURRENT FILING DATE: 2002-07-19  
PRIOR APPLICATION NUMBER: 60/306,161  
PRIOR FILING DATE: 2001-07-19  
PRIOR APPLICATION NUMBER: 60/306,150  
PRIOR FILING DATE: 2001-07-19  
PRIOR APPLICATION NUMBER: 60/331,477  
PRIOR FILING DATE: 2001-11-16  
NUMBER OF SEQ ID NOS: 125  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 69  
LENGTH: 1265  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-198-070-69

Query Match 3.6%; Score 103.5; DB 15; Length 1265;  
Best Local Similarity 19.0%; Pred. No. 2.2;  
Matches 114; Conservative 68; Mismatches 180; Indels 237; Gaps 30;

QY 30 STWLG---DRVITSTRTWALPTYNHLYKQI-----SSASTGA 65  
DB 49 STFLGLENEHLKKFQVTWEL--HNKHLFENVFSEPLLQSNLPALVSGIQLGTTTHTC 106  
QY 66 SNDNHYFGYSTPWGYFDENRFCHFSFRDW-----QRLNNWG- 104  
DB 107 SEDT---YSTLLQRYQRSEELRVAEWELECOKRIDAYVDEQMTKQRMLEDWEL 162  
QY 105 FRPKRLNFKLFNIQKVEVTNDGVTIANNLSTVQVFSDSHYQLP-----YVLGSAHQ 158  
DB 163 FKQRRF-----IEQLTNKKAVTGENNFDTMRHVLSSRLSMPCDNCYRRRCACD 214  
QY 159 GCL-----PPFPADV-----MIPOGYL-----TLNNGSOAVGRSSFYCL 194  
DB 215 DCSLSHILTCGIMDPVTDDIHQPLQVDPADYLAERSPPSVSSASGSSSPITI 274  
QY 195 EYFPSQMLR-TGNNTFSYTFEBV-PFHSSYASQSLDRLMPLIDQYLYLNKTQNSG 252  
DB 275 QOHPRLLITDSGAPTFCSDDDEDVAPLSAKFAD-----YPLSNYDDEV 319  
QY 253 SAQNKDLLFSGSPAGMSVQPKNMLPGPCYRQORVSKTKTDNNNSNFTWTGASKYNLNGR 312  
DB 320 VA-NMNGIHSELNGGEMALKDESP-----QISST----- 349  
QY 313 ESTINPGTAMASHKDEDEKFFPMGVMIFGKESAGASNTALJDNVMITDE---EIKATNP 369  
DB 350 -----SSSSSEADDEADGESSG-----EPPGAPKE--DGVLSRSPRTEESKADSP 394  
QY 370 VATERFGTVAVNFQSSSTDPAIGDVH-----AMGALP---GMWQDRDVTYL 412  
DB 395 PPS--YPT-----QQAEOAPNTCECHVCKQEAAGLTPSANTAGALPPGHQPLSPEKPTH- 446  
QY 413 QGPIWAKIPHTDGH--FH-----PSPLMGFGGLKNPPQILIKNTPVANPPA----- 458  
DB 447 --PALHLYPHIHGHVPLHTVPHLPLIHPTLYATPP-----FTHSKALPPAPVQNHNTN 498  
QY 459 -----EFSATKFASFTQYSTGQVSVIEIWELOKENSKRMP 495  
DB 499 KHQVFNASLQDHIYPSCFGNTPEWNSKFISSL-----WGSSEVMDKNWNP 543

RESULT 11  
US-10-108-605-103  
; Sequence 103, Application US/10108605  
; Publication No. US20020160934A1  
; GENERAL INFORMATION:  
; APPLICANT: Broadus, Julie  
; APPLICANT: Stam, Lynn

APPLICANT: Bachmann, Jane  
TITLE OF INVENTION: NUCLEIC ACID SEQUENCES FROM DROSOPHILA MELANOGASTER THAT ENCODE  
TITLE OF INVENTION: PROTEINS ESSENTIAL FOR LARVAL VIABILITY AND USES THEREOF  
FILE REFERENCE: 3113B  
CURRENT APPLICATION NUMBER: US/10/108,605  
CURRENT FILING DATE: 2002-03-27  
PRIOR APPLICATION NUMBER: US 09/761,142  
PRIOR FILING DATE: 2001-01-16  
PRIOR APPLICATION NUMBER: US 60/176,418  
PRIOR FILING DATE: 2000-01-14  
NUMBER OF SEQ ID NOS: 361  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 103  
LENGTH: 3712  
TYPE: PRT  
ORGANISM: Drosophila melanogaster  
US-10-108-605-103

Query Match 3.5%; Score 102; DB 15; Length 3712;  
Best Local Similarity 19.1%; Pred. No. 16;  
Matches 75; Conservative 55; Mismatches 163; Indels 100; Gaps 16;

QY 103 WGFPRKRLNFKL-----FNIOKVEVTNDGVTIANNLSTVQVFSSEYQLPYVLGSAH 157  
DB 3002 FGGHPRRLNTSISLQPNFDCCIDNVVINOQVVDLLEYVTG-----GGVE 3045  
QY 158 QGCLPPPADVFMIP-QYGYTLNNGSOAVGRSSFYCLEYF---PSQMLRTGNF----- 208  
DB 3046 EGCSAKFSTVVSYPHEYGFLMNVVS--SDNNLHVVLHFRTTQPNGLFYAANHDOSS 3102  
QY 209 TFSYTFEEVPEFHSSYASQSL--DRLMPLIDQYLYLNKTQNO---SSAQNKDLFS 262  
DB 3103 TIGLSIQDGLKLNSMSGQLVIDRIINDGED-HVTVQHTQGLRLTVDDVDNKRLL--- 3158  
QY 263 RGSFAGNSVQ-----PKNMLPGPCYRQORVSKTKTDNNNSNFTWTGASKY 307  
DB 3159 -GSPQPLLEGDIFFAGLPDNYRTPRNALASLAYFVGCI SDVTWNEEIIINFANSAEKN 3217  
QY 308 -NLNGRESINPGTAMASHKDEDEKFFPMGVMIFGKESAGASNTALJDNVMITDEEIIKA 366  
DB 3218 GNING-----CPHVLAYEPLSVPSYPSGDNEV---ESPMSNADTLPLPKDIESTLPP 3269  
QY 367 TNPVATERFGTVAVNFQSSSTDPAIGDVHAMGALPGMWQDRDVTYLQGIWAKIPHTDGH 426  
DB 3270 TTPPTTTTITTTTSTTTSTTTT-----TTPSPVIDEKEIEA----- 3310  
QY 427 FHPSPLMGFGGLKNPPQILIKNTPVANPPAE 459  
DB 3311 -----KTPQKILTTTPPAKINLPSD 3330

RESULT 12  
US-09-836-353A-145  
; Sequence 145, Application US/09836353A  
; Publication No. US20030129685A1  
; GENERAL INFORMATION:  
; APPLICANT: Ni et al.  
; TITLE OF INVENTION: 12 Human Secreted Proteins  
; FILE REFERENCE: PF489P1  
; CURRENT APPLICATION NUMBER: US/09/836,353A  
; CURRENT FILING DATE: 2001-04-18  
; PRIOR APPLICATION NUMBER: 60/198,407  
; PRIOR FILING DATE: 2000-04-19  
; PRIOR APPLICATION NUMBER: PCT/US99/25031  
; PRIOR FILING DATE: 1999-10-27  
; PRIOR APPLICATION NUMBER: 60/105,971  
; PRIOR FILING DATE: 1998-10-28  
; NUMBER OF SEQ ID NOS: 147  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 145  
; LENGTH: 439  
; TYPE: PRT





```

QY      394  ---VHAMGALPGMWVQODRDVYLQ-----PIWAKIPHTDGHFHPSPLM---GGF 436
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      907  YTLAVRDTGKEP--YTLQQLTLIEGLDNQPLPKLKITLKNKHVDAGAMRYELVKKNGEF 964
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      437  GLKNPPQILIKNTPVPANPPAEFSATKFAFITQYSTGQSV----- 480
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      965  RLHNPIKEQELRNDLVKAEQVERALIAQAELTTKKQKTEAKVRSKRAAFSDTPPDQQL 1024
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      481  --IEWELQKENSKRW-----NPEVQYTSNVA-----KSANVDF 511
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1025  NALQAELETINAQQQVAQAVQNKVTALNQNQNEQVKTQDKNLVLATALVEKETAQIDF 1084

```

```

      :|:|:|:|:|:|
Db      656 ELRFEDW-----TTSYATKQGREANPIEFHNCVEIETIYPAINNG 697

```

Search completed: July 17, 2003, 18:51:18

Job time : 27.7143 secs

```

RESULT 15
US-10-098-916A-11
; Sequence 11, Application US/10098916A
; Publication No. US20030096983A1
; GENERAL INFORMATION:
; APPLICANT: HECKEL, DAVID G.
; APPLICANT: GAHAN, LINDA J.
; TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A GENE CONFERRING RESISTANCE TO
; TITLE OF INVENTION: BACILLUS THURINGIENSIS TOXINS
; FILE REFERENCE: CXU-352
; CURRENT APPLICATION NUMBER: US/10/098,916A
; CURRENT FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: 60/276,180
; PRIOR FILING DATE: 2001-03-15
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 1714
; TYPE: PRT
; ORGANISM: Bombyx mori
US-10-098-916A-11

```

Query Match	3.5%;	Score 100.5;	DB 15;	Length 1714;	*
Best Local Similarity	19.1%;	Pred. No. 6.8;			
Matches 102;	Conservative 69;	Mismatches 184;	Indels 179;	Gaps 28;	

OY	92	PRDMORL-----INNNGFRPKRLNFKLENOYKEVTINDGVTTIANMLTSTVOYFSDSE	146
Db	235	PGEWIRMTWTVGIN-----EPLNFETNP LHSVLTALD--SLPNTHTVITLV-----	279
OY	147	YOLPYVLGSAHQCLPEFPADVEMIPOGYLTIN-----NGQAVGRSSFYCE----	195
Db	280	----QVENVEHR---PPRWEIFAVQOFDEKTAQSFPVR AIDGTGINKPIHYRLETAEE	332
OY	196	--YFPSQMLRTGNMFTFSY-----TFREEVPEHSS---YAHSQSL-----	229
Db	333	DTEFHRTIEGGRS GALLVPDIPDRDTLOREVPQLSITAYKYDNNESSATAANVIIVNDI	392
OY	230	-DRLMNP LIDQY-LYYLNRTQ-----NQSGAQNKDULLFSRGSPAGMSVQPKMWLPGPC-	281
Db	393	NDOQRP EPLFKEYRLNIMEETALT LNFDOEF GFHDRDL----GONAQYTVRL ES DVPADAA	448
OY	282	-----YRQQ-----RVSKTKTDN-NNSNFTWTGASK	306
Db	449	KAFYIAPEVG YQRQTFIMGTANH KMLDYEPBEQRI RL RVIATMDMNEEHVGA VAYYINL	508
OY	307	YNINGRESI INPGTAMASHKDEDKFFPMGSVMIFGKE-SAGASNTALDNVMITDEEIK	365
Db	509	INMNDEEPLFEHSVQNVSFKETEGGFVANAVBARHDRIDDRVEHTLMGN-----	558
OY	366	ATNPVATERFGTVAVNFOSSSTD PATGDVHAMGALPGMWQDR--DVYLQGP I WAKI PHT	423
Db	559	-----ANNYLSIDKD--TGDIH-----VTQDDPF DYHRQSEL FVQVRAD	595
OY	424	D---GHFHSP---LMGGFGLKNP PPQI-LIKNTP-VPANPRAEF SATKFASFITQYSTG	475
Db	596	DTLGEPHTATSOLLIHLEDINNTPTLR L PRGSPNVEENVPEGYI ITSEIRATDPDTTA	655
OY	476	QVSVELSEMELOKENSKRMNPEVOYT SNYAKSAN-VDF-----TVDNNG	517

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 2003, 18:28:05 ; Search time 19.1429 Seconds  
(without alignments)  
2681.720 Million cell updates/sec

Title: US-09-807-802A-17  
Perfect score: 2906  
Sequence: 1 MASGGAPMADNNEGADGVG.....NNGLYTEPRPIGTRYLTRPL 534

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_73:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2177.5	74.9%	504	1 VCPV3A	coat protein - ade
2	1676	57.7	732	2 S52210	coat protein VP1 -
3	517	17.8	673	1 VCPVB5	coat protein VP1 -
4	487.5	16.8	781	1 VCPV19	coat protein VP1 -
5	234.5	8.1	723	1 VCPVVP	coat protein VP1 -
6	232.5	8.0	729	1 VCPVNA	coat protein VP1 -
7	232.5	8.0	729	1 A60006	coat protein VP1 -
8	229	7.9	587	1 B44276	coat protein VP1 -
9	226	7.8	722	1 VCPV2	coat protein VP1 -
10	197	6.8	716	1 VCPV2M	coat protein VP1 -
11	196.5	6.8	718	1 VCPVIM	coat protein VP1 -
12	194	6.7	722	1 VCPVME	coat protein VP1 -
13	194	6.7	727	1 VCPV1F	coat protein VP1 -
14	189	6.5	584	2 S49594	capsid protein VP2
15	189	6.5	727	1 VCPVFP	coat protein VP1 -
16	188	6.5	737	1 VCPVCD	coat protein VP1 -
17	184	6.3	748	1 VCPVCP	coat protein VP1 -
18	177.5	6.1	702	1 VCPVAP	coat protein VP1 -
19	175	6.0	722	1 VCPVCN	coat protein VP1 -
20	123.5	4.2	690	2 AB0124	probable TonB-depe
21	122.5	4.2	880	1 SYBSVS	valine-tRNA ligase
22	121	4.2	635	2 F96660	protein F2K11.10 l
23	118	4.1	739	2 T52289	probable transkeo
24	115.5	4.0	345	1 G97024	probable phosphoe
25	113.5	3.9	2817	2 B97033	uncharacterized pr
26	109	3.8	418	2 T35753	probable periplasm
27	109	3.8	1072	2 A86827	hypothetical prote
28	109	3.8	1186	2 T18210	delta endotoxin -
29	108.5	3.7	1711	2 AB1283	peptidoglycan link

30	108	3.7	655	1 ALKRG	cyclomaldodextrin
31	108	3.7	772	1 A44052	outer layer protei
32	108	3.7	826	2 AD1683	penicillin-binding
33	107.5	3.7	413	2 T23098	hypothetical prote
34	107.5	3.7	648	2 S50856	whn protein - rat
35	107	3.7	777	2 AE0462	probable exported
36	107	3.7	1338	2 T30565	MAP kinase kinase
37	106.5	3.7	403	2 B87513	acyl-CoA dehydroge
38	106	3.6	1271	2 D64237	hypothetical prote
39	105.5	3.6	2500	2 G71609	hypothetical prote
40	105	3.6	1394	2 A29637	position-specific
41	104.5	3.6	559	2 S33724	transcription fact
42	104.5	3.6	717	2 H85171	DEF (Clal) protein
43	104	3.6	395	2 S50986	MAF1 protein - yea
44	104	3.6	715	2 H90977	hypothetical prote
45	104	3.6	719	2 F85824	hypothetical prote

ALIGNMENTS

RESULT 1		coat protein - adeno-associated virus type 2	
VCPV3A		C:Species: adeno-associated virus type 2	
		C:Date: 05-Apr-1983 #sequence_revision 05-Apr-1983 #text_change 16-Jul-1999	
		C:Accession: A03698	
		R:Sriwaslava, A.; Lusby, E.W.; Berns, K.I.	
		J. Virol. 45, 555-564, 1983	
		A>Title: Nucleotide sequence and organization of the adeno-associated virus 2 genome.	
		A:Reference number: A03694; MUID:83164299; PMID:6300419	
		A:Accession: A03698	
		A>Status: translation not shown	
		A:Molecule type: DNA	
		A:Residues: 1-504 <SRI>	
		A:Cross-references: EMBL:J01901; NID:g209616; PIDN:AAA42376.1; PID:g209621; EMBL:M12405;	
		C:Superfamily: adeno-associated virus coat protein	
		C:Keywords: coat protein	
Query Match		74.9%; Score 2177.5; DB 1; length 504;	
Best Local Similarity		82.5%; Pred. No. 2e-153;	
Matches 392; Conservative 33; Mismatches 47; Indels 3; Gaps 2;			
QY	1 MASGGAPMADNNEGADGVGNASGNWCHDSTWLGDRVITTSRTWALPTYNNHLYKQISS 60		
DB	1 MATSGAPMADNNEGADGVGNSSGNWCHDSTWMDRVITTSRTWALPTYNNHLYKQISS 60		
QY	61 ASTGASNDNHFGYSTPMWGYFDENRFFCHFSPRDQRLINNNGFRPKRLNFKLFNIQVK 120		
DB	61 QS-GASNDNHFGYSTPMWGYFDENRFFCHFSPRDQRLINNNGFRPKRLNFKLFNIQVK 119		
QY	121 EVTTNDGVTTIANNTSTVQVPSDSEYQLPYVLGSAHQGCLPFPADVEMIPOXYGLTLN 180		
DB	120 EVTQNDGTTIANNTSTVQVFTDSEYQLPYVLGSAHQGCLPFPADVEMVPOXYGLTLN 179		
QY	181 NGSQAVGRSSFYCLEYFPSSQMLRTGNNFTFSYTFEEVPPHSSYAHQSQSLDRMLNPLIDQY 240		
DB	180 NGSQAVGRSSFYCLEYFPSSQMLRTGNNFTFSYTFEDVPFHSSYAHQSQSLDRMLNPLIDQY 239		
QY	241 LYYLNRTONQSGSAQNKDLFSRSPAGMSVQPKMWLPGPCYRQQRVSKTKTDNNNSNFT 300		
DB	240 LYYLSRNTPSGTTTQSRLOFSQASDIRDQSRNWLPGPCYRQQRVSKTSADNNNSSEYS 299		
QY	301 WTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMGSGVMIFGKESAGASNTALDNVMTID 360		
DB	300 WTGATKYHLNGRDSLVP--AMASHKDEDEKFFPQSGVLLFGKQSEKTNVNIKVMITD 357		
QY	361 EEEIKATNPVATERFGTAAVNFQSSSTDPAAGDVHAMGALPGMWQDRDVLQGPIMAKI 420		
DB	358 EEEIGTNPVATEQYGSVSTNLQRGNQAATADVNTQGVLPQMWQDRDVLQGPIMAKI 417		
QY	421 PHTDGHFSPPLMGFGKLNPPQILLIKNTVPANPPAFSATKFAFITQYSTG 475		



Db 418 PHTDGHFHPSPLMGFGFLKHPPOILLIKNTVPVNPANPSTTFSAKFAFITQYSTG 472

RESULT 2

S52210

coat protein VP1 - muscovy duck parvovirus

N/Alternate names: VP1 protein

C/Species: muscovy duck parvovirus

C/Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 17-Mar-2000

C/Accession: S52210

R/Zadori, Z.; Erdei, J.; Nagy, J.; Kisari, J.

submitted to the EMBL Data Library, September 1993

A/Reference number: S52209

A/Accession: S52210

A/Molecule type: DNA

A/Residues: 1-732 <ZAD>

A/Cross-references: EMBL:X75093; NID:g609091; PIDN:CA52984.1; PID:g609093

A/Experimental source: strain FM

C/Genetics:

A/Gene: VP1

C/Superfamily: parvovirus coat protein

C/Keywords: coat protein

Query Match 57.7%; Score 1676; DB 2; Length 732;

Best Local Similarity 56.9%; Pred. No. 4.8e-116;

Matches 308; Conservative 77; Mismatches 142; Indels 14; Gaps 6;

OY 1 MASGGAPMADNNEGADGVGNASGNWHCDSTWLGDRIITSTRTWALPTYNNHLYKOISS 60

Db 199 MAEGSGAMGDSAGADGVGNASGNWHCDSTWLGDRIITSTRTWALPTYNNHLYQAITS 258

OY 61 ASTGASNDNHFGYSTPWCYFDFENRHFCHFSPRDWORLNNNWGFRPKRLNEKLFENIQVK 120

Db 259 GTNPDSN-TQYAGYSTPWCYFDFENRHFCHFSPRDWORLNNHWGIRPKALKKIFNVQVK 317

OY 121 EVTTNDGVTIANNLTSIVQVFSSEYQLPYVLGSAHQGLPPFPADVEMIPQYGYLTIN 180

Db 318 EVTTQDOTKTIANNLTSITQIFTDNEHQLPYVLGSAHQGLPPFPADVEMIPQYGYCTMH 377

OY 181 ---NGSQAVGRSSFYCLEYFPSSQMLRTGNNFTFSYTFEEVPPHSSYAHQSQDLRLMPLI 237

Db 378 TNQSGARFNDRSAFYCLEYFPSSQMLRTGNNFEFSFEFEVPPHSMFAHSQDLRLMPLI 437

OY 238 DQYLYLNRTONQSGAQNKDLLFSRGSAGMSVQPKNWLPGPCYRQOVSQTK--TDNN 295

Db 438 DQYLMNFSSEV-NGGRNAQ-----FKKAVKGAFGAGRNWLPQPKLLDQRYAVYSGGTDNY 491

OY 296 NSNFTWTGASKYNLNGRESIINPGTAMASHKODEDKFFPMGVMIFGKE--SAGASNTAL 353

Db 492 ANWSIWSKGNKVFLLKDRYLLQPGVATHTHTEDQSSVPAQNIIGIAKDPYRSGSTLAGI 551

OY 354 DNVMITDEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMGALPGMWQDRDVILO 413

Db 552 SDIMVTDEQEIAPTNGVWRPYGLTVTNEQNTTAPTNAELJVLGALPGMWQNRDIYLO 611

OY 414 GPIWAKIPHTDGHFHPSPLMGFGFLKHPPOILLIKNTVPVNPANPAPAFSATKFAFITQYS 473

Db 612 GPIWAKIPKTDGKHPSPNLGGFLHNPFPQVFIKNTVPVNPADPPLLEYVNGKNSYITQYS 671

OY 474 TGQVSVEIEMELQKENSKRWNPEVOYTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRP 533

Db 672 TGQCTVEMWELKRENSKRWNPEIOFTSNFGNRTSTMFAPNETGGYVEDRLIGTRYLTQN 731

OY 534 L 534

Db 732 L 732

RESULT 3

VCPVB5

coat protein VP1 - bovine parvovirus

N/Contains: coat protein VP2

C/Species: bovine parvovirus

C/Date: 28-Dec-1987 #sequence\_revision 28-Dec-1987 #text\_change 16-Jul-1999

C/Accession: A26104

R/Chen, K.C.; Shull, B.C.; Moses, E.A.; Lederman, M.; Stout, E.R.; Bates, R.C.

J. Virol. 60, 1085-1097, 1986

A/Title: Complete nucleotide sequence and genome organization of bovine parvovirus.

A/Reference number: A26104; MUID:87061184; PMID:3783814

A/Accession: A26104

A/Molecule type: DNA

A/Residues: 1-673 <CHE>

A/Cross-references: EMBL:M14363; NID:g333454; PIDN:AAB59847.1; PID:g808805

C/Superfamily: parvovirus coat protein

C/Keywords: coat protein

F/136-673/Product: coat protein VP2 #status predicted <VP2>

Query Match 17.8%; Score 517; DB 1; Length 673;

Best Local Similarity 25.1%; Pred. No. 2.9e-30;

Matches 143; Conservative 85; Mismatches 208; Indels 134; Gaps 17;

OY 5 GGAPMADNNEGADGVGNASGNWHCDSTWLGDRIITSTRTWALPTYNNHLYKOISSASTG 64

Db 163 GSGSVGGGRGSGVGYSTGWTGTIFSENIIVTKNTROFICDIKNGHLYKS-EVLNTG 221

OY 65 ASNDNHFGYSTPWCYFDFENRHFCHFSPRDWORLNNNWGFRPKRLNEKLFENIQVEVTT 124

Db 222 DTAHQY-AITPWSYFNFQYSSHSFENDWQHLVNDYERFRKAMIVRVNLIQIKQIMT 280

OY 125 NDGVTIANN-LTSTVQVFSSEYQLPYVLGSAHQGLPPFPADVEMIPQYGYLTLL---- 179

Db 281 DGAMGYVANNDLTAGMHIFCDGHRYPVQHPWDQCMPELPSIWELLPOYAYIPAPISV 340

OY 180 --NNGSQAVGRS-----SFYCLEYFPSSQMLRTGNNFTFSYTFEEVPPHSSYAHQSQDLRL 232

Db 341 VDNNTTIVYEEHLKGVPLMYLENDEHVLRLNG-----RI 375

OY 233 MNPLIDQYLYLNRTON-----QSGSAQNKDLLFSRGS-----AGMSVQPKNWL 277

Db 376 YRIYIQLWRLMDRKQHHIQHASDVOSTGQKQKLLIQRTQPNKQRFQNALRTSNM 435

OY 278 PGPCYRQOVSQTKTDNNNSNFTWTGASKYNLNGRESIINPGTAMASHKODEDKFFPMG 337

Db 436 SGP-----GIARGTNATLQTSAGALVTMTNGAD----VSG 469

OY 338 VM-----IFGKE-----SAGASNTALDNVMITDEEIKATNPVAT 372

Db 470 VRAVRVGYSTDPYIGQQPESDLLRLYSASAAEQQNPILEN-----AAR 515

OY 373 ERFGTAVNFQSSSTDPATGDVHAMGALPGMWQDRDVILOGPIWAKIPHTDGHFHPSP 432

Db 516 HTFTREARFKLITGSGADGDYKEMWMLPNQMWDSAPISRYNPIWVKVPRVNRKTLDTQ 575

OY 433 MGGFGLKNPPOILLIK--NTPVNPANPAPAFSATKFAFITQYSTQVSEIEMELQKENS 490

Db 576 DGSIPMSHPGTIFIKLARIPVPGND-----SFLNIYVTQVSCVEVWEVEKRG 626

OY 491 KRWNPVOYTSNYAKSANVD-FTVDNNGLY 519

Db 627 KNWRPEYMS--ATNMSVDAYTINNAGVY 653

RESULT 4

VCPV19

coat protein VP1 - parvovirus B19 (strain Au)

C/Species: parvovirus B19

A/Note: host Homo sapiens (man)

C/Date: 28-Dec-1987 #sequence\_revision 28-Dec-1987 #text\_change 16-Jul-1999

C/Accession: A24299

R/Shade, R.O.; Blundell, M.C.; Cotmore, S.F.; Tattersall, P.; Astell, C.R.

J. Virol. 58, 921-936, 1986

A/Title: Nucleotide sequence and genome organization of human parvovirus B19 isolated fr

A/Reference number: A24299; MUID:86200451; PMID:3701931

A/Accession: A24299

A/Molecule type: DNA

A/Residues: 1-781 <SHA>



```
QY 3 SGGAPMADNNEGADGVGNASG--NMWCHDSITWLGCD---RVITSTRTWALPTYNNHLYKQ 57
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 174 SGGGGG-GGGGRGAGGVSTGTENNQTEFOYLGELVRITAHASRLIHLNMPHEHETTKR 232
QY 58 IS--SASTGAS-----NDNHYEGYSTPWGYFDENRFCHFSPRDWORLNNNWGFRPKRLN 111
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 233 IHLNSESAGVAGQVDDAHTQMTWPSLIDANAMGWENPADWQOLISNNTEINLVSE 292
QY 112 FKLFINIQKEVT---TNDGVTTIANLITSTVOVFSDEYQOLPYVLGSAHQGLPPPADV 168
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 293 QEIFNVVLKTTITESATSPPTKIYNNDLTASLMAVADTNNLTPTTPAPRSETLGFYPWLP 352
QY 169 FMIPQGYL-----TLNNGSQAVGRS-----SPYCLE-YFPSQMLRTGNMF 208
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 353 TKPTQYRYLSCIRNLNPTTYTGQSQOITDSIQGLHSDIMEFTYIENAVPIHLRTGDEF 412
QY 209 -TFSYTFEEVPHSSSYAHSQSLDR-----LMNPLI--DOYLYLNRTQNGSGSAQNK 257
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 413 STGIYHFDTKPL--KLTHSWQTNRSGLPKLLEPTTEGDQHPGTLPAANTRKGYHQT 470
QY 258 DLLFSRGSBAGMSVQPKNWLPGPCYRQQRVSKTKTDNNNSNFTWTGASKYNLNGRESIIN 317
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 471 NNSYTEAT---AIRP-----AQVGYNTPYMNFESYNGGPF-----LT 504
QY 318 P--GTAMASHKODEDEKFFPMGSMIFGKESAGASNTALD--NMITDEEEK--ATNP- 369
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 505 PIVPTADTQYNDDEPN-----GAIRFTMDYOHGLTTSQELERYTFNPQ 549
QY 370 -----VATERFGIVA-VNFQSSS-----TDPATG--DVHAMGALP-----GMVWQDRDY 410
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 550 SKCGRAPKQQFNOQAPLNLENTNNGTLTPSDPIGKSNMHEMNTLNTYGPALTALNTAPV 609
QY 411 YLQGPIMAKIPHTD--GHFHPSPLMGFGFLK-NPPQILIKNTVPVAPNPAPFSA-TKFA 466
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 610 FPNQIWDKELDTLKPRLH--VTAPFVCKNPPGQLFVKIAP--NLTDENADSPQQ 663
QY 467 SFITQYSTGVSEIEMELQKENSKRWNPEVOYTSNYAKSANVDFTVNNGLYTEPRPIG 526
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 664 PRITTSNFWWKGTLTFTAKMRSSNMWNP1QOHTT-----TAENIGNYI-PTNIG 712
```

## RESULT 7

```
A60006
coat protein VP1 - porcine parvovirus (strain 90HS)
N/Contains: coat protein VP2
C/Species: porcine parvovirus
C/Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 30-Sep-1993
C/Accession: A60006
R/Sakurai, M.; Nishimori, T.; Ushimi, C.; Nakajima, H.
Virus Res. 13, 79-86, 1989
A/Title: Nucleotide sequence of capsid protein gene of porcine parvovirus.
A/Reference number: A60006; MUID:89319168; PMID:2750278
A/Accession: A60006
A/Molecule type: DNA
A/Residues: 1-729 <SAK>
C/Superfamily: parvovirus coat protein
C/Keywords: coat protein; glycoprotein
F/151-729/Product: coat protein VP2 #status predicted <VP2>
F/172,198,282,330,433,471,573,604,651/Binding site: carbohydrate (Asn) (covalent) #statu
```

Query Match 8.0%; Score 232.5; DB 1; Length 729;  
Best Local Similarity 23.0%; Pred. No. 3.8e-09;  
Matches 137; Conservative 86; Mismatches 245; Indels 127; Gaps 31;

```
QY 3 SGGAPMADNNEGADGVGNASGNMH--CDSTWLGCD---RVITSTRTWALPTYNNHLYKQ 57
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 174 SGGGGG-GGGGRGAGGVSTGTFNNQTEFOYLGELVRITAHASRLIHLNMPHEHETTKR 232
QY 58 IS--SASTGAS-----NDNHYFGYSTPWGYFDENRFCHFSPRDWORLNNNWGFRPKRLN 111
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 233 IHLNSESAGVAGQVDDAHTQMTWPSLIDANAMGWENPADWQOLISNNTEINLVSE 292
QY 112 FKLFINIQKEVT---TNDGVTTIANLITSTVOVFSDEYQOLPYVLGSAHQGLPPPADV 168
```

```
Db 293 QEIFNVVLKTTITESATSPPTKIYNNDLTASLMAVADTNNLTPTTPAPRSETLGFYPWLP 352
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 169 FMIPQGYL-----TLNNGSQAVGRS-----SPYCLE-YFPSQMLRTGNMF 208
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 353 TKPTQYRYLSCIRNLNPTTYTGQSQOITDSIQGLHSDIMEFTYIENAVPIHLRTGDEF 412
QY 209 -TFSYTFEEVPHSSSYAHSQSLDR-----LMNPLI--DOYLYLNRTQNGSGSAQNK 257
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 413 STGIYHFDTKPL--KLTHSWQTNRSGLPKLLEPTTEGDQHPGTLPAANTRKGYHQT 470
QY 258 DLLFSRGSBAGMSVQPKNWLPGPCYRQQRVSKTKTDNNNSNFTWTGASKYNLNGRESIIN 317
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 471 NNSYTEAT---AIRP-----AQVGYNTPYMNFESYNGGPF-----LT 504
QY 318 P--GTAMASHKODEDEKFFPMGSMIFGKESAGASNTALDVMITDEEEKATNP----- 369
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 505 PIVPTADTQYNDDEPN-----PNGAIRFTMGYOHGLTTS-----SQELERYTFNPQSKGR 554
QY 370 VATERFGIVA-VNFQSSS-----TDPATG--DVHAMGALP-----GMVWQDRDYLOCP 415
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 555 APKQQFNOQAPLNLENTNNGTLTPSDPIGKSNMHEMNTLNTYGPALTALNTAPVFPNGQ 614
QY 416 IWAKIPHTD--GHFHPSPLMGFGFLK-NPPQILIKNTVPVAPNPAPFSA-TKFAFITQ 471
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 615 IWDKELDTLKPRLH--VTAPFVCKNPPGQLFVKIAP--NLTDENADSPQQPRIIT 668
QY 472 YSTGVSEIEMELQKENSKRWNPEVOYTSNYAKSANVDFTVNNGLYTEPRPIG 526
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 669 YSNFWWKGTLTFTAKMRSSNMWNP1QOHTT-----TAENIGNYI-PTNIG 712
```

## RESULT 8

```
B44276
coat protein VP1 - parvovirus LuIII.
C/Species: parvovirus LuIII
C/Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 26-Feb-1999
C/Accession: B44276
R/Diffoot, N.; Chen, K.C.; Bates, R.C.; Lederma, M.
Virology 192, 339-345, 1993
A/Title: The complete nucleotide sequence of parvovirus LuIII and localization of a unique
A/Reference number: A44276; MUID:93297126; PMID:8517025
A/Accession: B44276
A/Status: translation not shown
A/Molecule type: DNA
A/Residues: 1-587 <DIR>
A/Cross-references: GB:M81888
C/Superfamily: parvovirus coat protein
C/Keywords: coat protein; glycoprotein
F/49,90,220,304,371,503,511,514,539,571/Binding site: carbohydrate (Asn) (covalent) #statu
```

Query Match 7.9%; Score 229; DB 1; Length 587;  
Best Local Similarity 22.7%; Pred. No. 5e-09;  
Matches 133; Conservative 74; Mismatches 228; Indels 150; Gaps 28;

```
QY 3 SGGAPMADNNEGADGVGNASGNMHCD--TWLGD---RVITSTRTWAL--PTYNN--- 52
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 30 SGGG-----GSGGGGVGVSTGTYNQTHYKFLGDGWVEITAYSTRMVLNMPKSENYCR 83
QY 53 ---HLYKQISSASTGASNDNHYFGYSTPWGYFDENRFCHFSPRDWORLNNNWGFRPKR 109
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 84 VRVHNTNDGTASHMAMDADHEQIW-TFWSLV DANAMGWFPQSDWQYISNNMIHINLHS 142
QY 110 LNFKLFINIQKEVT-TNDGVTTIA---NNLTSTVOVFSDEYQOLPYVLGSAHQGL---- 161
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 143 LDQELFNVAIKVTQNTGAELIKYNNNDLTAAMVALDSNNILPYTPAIDNQETLGFYP 202
QY 162 --PPFPA-----DVFMIPQGYLTLNNGSQAVGRSSFYCLEYFPS-QMLRTGN 206
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 203 WKPTIPSPRYFYSCDRNLSTVYKDEAGTITDTMGLASGLNSQEFFTIENTQRIINLRTGD 262
QY 207 NF-TFSYTFEEVPHSSSYAHSQSLDRMLNPLIDQ-----YLYLNRTQN- 249
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
```



```
Db 263 EYATGTYVFDTEPIRLTHWTQNRHLGQPPQITELPSSDTANATLTARGYRSGLTQIQGR 322
Qy 250 -----QSGSAQNKD-LLFSRGSFAGMSVQPKNWLPG-----PCYRQQR 286
Db 323 NDVTEATRVRPAYVFCQPHDNFETSIRAGPFKYVVPVADITQGLDHANGSLRITYDKQH 382
Qy 287 VSKTKTDNNNSNFTWTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMGSMVIFGKESA 346
Db 383 GQSWASQNNKDRYTW-DAVNYD-SGR----- 406
Qy 347 GASNTALDNVMTLDEEIKATNPVATER---FGTAVNFGS--SSTDPATGDVHAMGALP 401
Db 407 WTNNCFIQSVFTSEPN---ANQILTNRDLNLAGKTDIHFTNAFNSYGLTAPFH----- 457
Qy 402 GMVWQDRDVLVLOGPIWAK---IPHTDGHFHPSPLMGFGGLKNPPQILIKNTP--VPANP 456
Db 458 -----PAPIYQGOIWDKELDLHKPRLHTQAPFV--CKNNAPGQLLVRLAPNLTQYD 509
Qy 457 PAEFSATKPAASFITQYSTGQVSVEIEMWLOKNSKRMNPEVOYTS 501
Db 510 PNSSNLSRIYTYGTFFWKGLTLKAK--MRPNA-TWNPVFQISA 550
```

## RESULT 9

```
VCPV2
coat protein VP1 - parvovirus H1
C:Species: parvovirus H1
A:Note: host Homo sapiens (man)
C:Date: 14-Nov-1983 #sequence_revision 14-Nov-1983 #text_change 08-Apr-1994
C:Accession: A03699
R:Rhode III, S.L.; Paradiso, P.R.
J. Virol. 45, 173-184, 1983
A:Title: Parvovirus genome: nucleotide sequence of H-1 and mapping of its genes by hybridization
A:Reference number: A03695; MUID:83112183; PMID:6823009
A:Accession: A03699
A:Molecule type: DNA
A:Residues: 1-722 <RHO>
A:Cross-references: EMBL:X01457; EMBL:J02198
C:Superfamily: parvovirus coat protein
C:Keywords: coat protein
```

```
Query Match 7.8%; Score 226; DB 1; Length 722;
Best Local Similarity 21.7%; Pred. No. 1.1e-08;
Matches 128; Conservative 83; Mismatches 234; Indels 146; Gaps 26;

Qy 2 ASGGAPMADNNEGADGVGNAGNWHCDSTW--LGDRVITSTRTWAL----- 47
Db 153 ADGGGS-SCGGGSGGGGIGVSTGYDNGQTYKFLGDGWVEITAHASRLHLGMPSENYC 211
Qy 48 --PTVNNHLKYQISSASTGASNDNHFGYSTPWGYFDNRFHCHFSRDMQRLINNWGF 105
Db 212 RVTVHNNQTTGHGTQKVGKGNMAYDTHQIW--TPWSLVDANAMGWVFPQSDWQFQNSMESL 270
Qy 106 RPKRLNFKLNIQVKEVT-----NDGYTTIANNLSTVQVFSDEYQLPYVLGSAHQGC 160
Db 271 NLDSLQGELEFNVVKTVEQCGAGQDAIKYNNDLTACMMVALDSNNILPYTPAQTSET 330
Qy 161 L-----PPFPAD--VEMIPQYGYLTINNGSQ-----AVG-----RSSFYCLE-YFPS 199
Db 331 LGFYPMKPTAPAPRYRYFFMPKRLSVTSSNSAEGTQITDTIGEPQALNSQFTIENTLPI 390
Qy 200 QMLRTGNFTF-SYTFEEVFFHSSYA-----HSQSLDRLMN-PLIDQYLYLNRTQNSG 252
Db 391 TLLRTGDEFTTGTIYFNTDPLKLTHTWQNRHLACLQGITDLPSTDTATASLTANGDRFG 450
Qy 253 SAQNKDLF-----SRGSPAGMSVQPKNWLPGPCYRQQRVSK 289
Db 451 STQTQNVNVTALRTRPAQIGFMQPHDNFEANRGPFKVPVP-----LDI 497
Qy 290 TKTDNNNSNFTWTGASKYNLNGR--ESLIINPGTA-----MASHKDEDEKFFPMG 337
Db 498 TAGEDHDAN-----GAIRFNYSKQHGEDMAKQGAAPERYTWDAIDSAAGRTARCFV----- 549
```

```
Qy 338 VMIFGESAGASNTALDNVMTLDEEIIKA-TNPVATERFGTAVNFGSSSTDPATGDVHA 396
Db 550 -----QSAFISIPPNOQLQREDAIAGRTNMHYTNVFNYSYGPLSAFPHDP----- 596
Qy 397 MGALPGMWQDRDVLVLOGPIWAK---IPHTDGHFHPSPLMGFGGLKNPPQILIKNTPVP 453
Db 597 -----IYPNGQIWDKELDLHKPRLHVTAPFV---CKNNPPGQLFVHLGP-- 638
Qy 454 ANPPAEF--SATKPAFITQYSTGQVSVEIEMWLOKNSKRMNPEVOYTSN 502
Db 639 -NLTDQDPNSTTVSRIVT-YSTFYWKGLLKFKAKLRLPLTNWNPVYQATTD 687
```

## RESULT 10

```
VCPV2M
coat protein VP1 - minute virus of mice
C:Species: minute virus of mice, murine parvovirus
C:Date: 14-Nov-1983 #sequence_revision 14-Nov-1983 #text_change 08-Apr-1994
C:Accession: A03700
R:Atell, C.R.; Thomson, M.; Merchlinsky, M.; Ward, D.C.
Nucleic Acids Res. 11, 999-1018, 1983
A:Title: The complete DNA sequence of minute virus of mice, an autonomous parvovirus.
A:Reference number: A03696; MUID:83143341; PMID:6298737
A:Accession: A03700
A:Molecule type: DNA
A:Residues: 1-716 <AST>
A:Cross-references: EMBL:V01115
C:Superfamily: parvovirus coat protein
C:Keywords: coat protein
```

```
Query Match 6.8%; Score 197; DB 1; Length 716;
Best Local Similarity 20.0%; Pred. No. 1.6e-06;
Matches 126; Conservative 88; Mismatches 241; Indels 174; Gaps 28;
```

```
Qy 3 SGGAPMADNNEGADGVGNAGNWHCDSTW--TWLGD--RVITSTRTWALPTYNNHLKYQ 57
Db 160 SGG-----SGGGGVGVSTGYDNGQTHYRFLGDGWVEITATLRLVHLNMPKSENYC 213
Qy 58 I-----SSASTGASNDNHFGYSTPWGYFDNRFHCHFSRDMQRLINNWGFPRKRL 110
Db 214 IRVHNTDTSVKGNMAKDAHEQIWTWPSLVDANAMGWVLPQSDWQYICNTMSQNLVSL 273
Qy 111 NFKLNIQVKEVTND---GVTTIANNLSTVQVFSDEYQLPYVLGSAHQGLPPFPA 166
Db 274 DQELFNVLKVTVEQDLGGQAIKIYNNDLTACMVAVDSNNILPYTPAANSMETLGFYFW 333
Qy 167 DVEMIPQYGY-----LTINNGSQAV-----GRSIFYCLEYFPS-QMLRTGN 207
Db 334 KPTIASPYRYFPCVDRDLSTVYENQEGTVEHNWMTGTPKGIPOFTIENTQOITLLRTGDE 393
Qy 208 F-TFSYTFEEVFFHSSYASQSLDRLMN--PLIDQYLYLNRTQNSGSAQNKDLFSGR 264
Db 394 FATGYTFYFDTSV--KLTHWTQNRQLGQPLLSTF---PEADTDAGT-----LTAQG 441
Qy 265 SPAGMSVQPKNWL-----DGPCYRQQRVSKTKTDNNNSNFT 300
Db 442 SRHGTQMGVNVWVSEAIRTRPAQVFCQPHDNFEASRAGP-FAAPKVPADITQGVDEKAN 500
Qy 301 WTGASKYNLNGRESIINPG-----TAMASHKDEDEKFF-----PMGVMIF 341
Db 501 GSVRYSYGKHGEMWASHGAPABERYTWDETSFGSGRDTKDGFIQSAPLVVPPLNGI--- 557
Qy 342 GKESAGASNTALDNVMTLDEEIKATNPVATERFGTAVNFGS--SSTDPATGDVHAMGA 399
Db 558 -----LTNANPIGTKN---DIHFSNVFNSYGPLTAFSH----- 587
Qy 400 LPGMWQDRDVLVLOGPIWAK---IPHTDGHFHPSPLMGFGGLKNPPQILIKNTP--VPA 454
Db 588 -----PSPVYQGOIWDKELDLHKPRLHTAPFV---CKNNAPGQLVRLGPLNLTQ 637
Qy 455 NPPEFSATKPAASFITQYSTGQVSVEIEMWLOKNSKRMNPEVOYTSNYAKSANVDFTV 514
Db 638 YDENGATLSRIYTYGTFFWKGLTMRAKLRA---NTTWNPVYQVSAB-----D 682
```

```
QY      515  NNGLY---TEPRPIGT-----RYLTRPL 534
          |   |   |   |   |   |   |   |
Db      683  NGNSYMSVTKMLPTATGNMQSVPLTRPV 711
```

## RESULT 11

coat protein VP1 - minute virus of mice (strain MMV1)  
C/Species: minute virus of mice, murine parvovirus  
C/Date: 28-Dec-1987 #sequence\_revision 28-Dec-1987 #text\_change 28-Jul-2000  
C/Accession: B23008; B29510  
R/Sahl, R.; McMaster, G.K.; Hirt, B.  
Nucleic Acids Res. 13, 3617-3633, 1985  
A/Title: DNA sequence comparison between two tissue-specific variants of the autonomous  
A/Reference number: A23008; MUID:95242059; PMID:3855242  
A/Accession: B23008  
A/Molecule type: DNA  
A/Residues: 1-718 <SAH>  
A/Cross-references: EMBL:X02481; NID:g60918; PIDN:CAB46507.1; PID:g5419928  
R/Astell, C.R.; Gardiner, E.M.; Tattersall, P.  
J. Virol. 57, 656-669, 1986  
A/Title: DNA sequence of the lymphotropic variant of minute virus of mice, MMV(i), and  
A/Reference number: A29510; MUID:86115415; PMID:3502703  
A/Accession: B29510  
A/Molecule type: DNA  
A/Residues: 1-143, 'A', 145-718 <AST>  
A/Cross-references: EMBL:M12032  
C/Superfamily: parvovirus coat protein  
C/Keywords: coat protein

Query Match

Query Match	6.8%;	Score 196.5;	DB 1;	Length 718;
Best Local Similarity	20.5%;	Pred. No. 1.7e-06;		
Matches 122;	Conservative 82;	Mismatches 243;	Indels 147;	Gaps 25

```

QY      3 SGGGAPMADNNEGADVGNASGNWHCHDS--TWLGD---RYITTTSTRTVALPTYNHLYKQ   57
Db     161 SGGG-----GSGGGGVSTGSYDNQTHYRFLGDGWVEITALATRLVHLNMPKSENYCR   214
QY      58 I-----SSASTGASNNDNHFGYSTPWGYFDENRFHCHESPRDWORLINNNWGFRPKRL   110
Db     215 IRVHNNTDTSVKGNMAKDPAHQIWTWPMSLV DANAWGWIQPSDWQYICNTMSQLNLVSL    274
QY     111 NFKLFNIQVKEVTIND---GVTTIANLLSTVOVFSDEXULPYVLGSAHQCLPFPFA   166
Db     275 DQEIFNVVLKTVTEQDSGGQAIRKYNNDLTACMMVAVDNNILPYTPAANSMETLGFFPW   334
QY     167 DVEMIPQXY-----LTNN-----GSOAVGRSSFYCLEYFPS-QMLRTGN   206
Db     335 KPTIASPRYYFCVDRDLSTVIYENOEGTIEHNVMGT PKGMNSQF TIENTOQTILLRTGD   394
QY     207 NF-TESYTPEEVPPFHSSIAHSOSLDRLMN--PLIDQYL YLNRTONQS GSAQNKLLESR   263
Db     395 EFATGTYFFDTNPV--KLTHWQTRQLGQPPLLSTF----PEADTDAGT-----LTAQ   442
QY     264 GSPAGMSVQPKNWL-----PBCYRQRQVRVSKTKTDNNNSNF   299
Db     443 GSRHGATOMEVNWNVSEAIRTRPAQVGFQCOPHNDFEASRAGP-FAAPKV PADVTQGVDREA   501
QY     300 TWTGASKYNLNGRESIINPGTAMASHKDEDKF-----FPMGVMIFGESAGASN   350
Db     502 NGSVRYSYGKHGENNAHGAAPERITYTWDETNFGSGRDTRDGFIQSA PLVPPLINGILT   561
QY     351 TALDNVMTDEEIKATNPVATERFGTVAVNFQS--SSTD PATGDVHAMGALPGMWQDR   408
Db     562 NA-----NFIGTKN---DIHFSNVENSIGPULTAFSH-----PS   591
QY     409 DVLLOGPIMAK--IPTHDGHFHP SPLMGFGGLKNPPQIILIKNTP--VPANPAAEF SAT   463
Db     592 PVPPOGOIWDKELDELHKRHLITAPFV--CKNNAPGOMLVRLGPLNLTDOYPDNGATLS   648
QY     464 KEASFITOYSTGOVSVEIEMELOKENS KRWNPEVOYTSNYAKSANVDFTVDNNG   517

```

Db 649 RIVTYGTFFWKGLTRAKLRA-----NTWNPVYQ-----VSVEDNG 686

## RESULT 12

coat protein VP1 - mink enteritis virus (strain Abashiri)  
N/Contains: coat protein VP2  
C/Species: mink enteritis virus, MEV  
C/Date: 30-Jun-1992 #sequence\_rev1stion 30-Jun-1992 #text\_change 12-Apr-1996  
C/Accession: B38350  
R/Karjatsumari, T.; Horiuchi, M.; Hama, E.; Yaguchi, K.; Ishiguro, N.; Goto, H.; Shinaga  
J. Gen. Virol. 72, 867-875, 1991  
A>Title: Construction and nucleotide sequence analysis of an infectious DNA clone of the  
A/Reference number: A38350; MUID:91202123; PMID:2016597  
A/Accession: B38350  
A/Molecule type: DNA  
A/Residues: 1-722 <KAR>  
A/Cross-references: GB:D00765  
C/Superfamily: parvovirus coat protein  
C/Keywords: coat protein  
F/139-722/Product: coat protein VP2 #status predicted <VP2>

Query Match

Query Match	6.7%;	Score 194;	DB 1;	Length 722;
Best Local Similarity	20.7%;	Pred. No. 2.7e-06;		
Matches 136;	Conservative 73;	Mismatches 215;	Indels 232;	Gaps 32;

QY	5	GAAPMADNNEGADGVGNASGNHCDSTWLGDRLVITTTST-----RTWALDPTYNN-	52
Db	148	GGOP-AVNERATGSGNGSGGGGGGS--GGVGISTGTENNQTETFKFLENGWEITANSS	204
QY	53	---HL-----YKQI-----SSASTGASNDNHFGYSTPWGCFDFNRFHCHSPRDW	95
Db	205	RLVHNNMPESENYKRVVNVNNDKTAVKGNMALLDTHVQIVTFPWSLVDANAMGVWENPGDW	264
QY	96	QRLINNNGWGRPKRLNFKLFNIQVKEVT--TNDGVTTIANNLSTVQVFSDEXQLPYV	152
Db	265	QLIVNTMSELHLVSFEQEIFNVVLKTVSESATQPTKYVNNDLTASLMVALDSNNTMPT	324
QY	153	LGSAHQGL-----PPFPA-----DVFMIPQYGLTLNNGSQAVGRS-----	189
Db	325	PAAMSETLGFYPWKFTIPTWRMYFQWDRTLIPSH-----TGTSGTPTNIYHGTDPDD	378
QY	190	-SFYCLE-YFPQMLRTGNF--TF-----SYTPE-----VPFHSYAHSQSL	229
Db	379	WQFYTIENSVPVHLRTGDEFATGTFFDCKPCRLTHTWQTNRALGLPPLNSLEQSEGA	438
QY	230	DR-----LMNPLIDQYL--YYLNRTQNGS-----	251
Db	439	TNFGDIGVQDKRRGVTQMGNTDYITEATIMRPAEVGSAPYYSFEASTQGPXTPIAAG	498
QY	252	-GSAQNKDLLFSRGSFAGMSVQPKNMLPGPCY--RQQRVSKTKTDNNSNFTWTGASKY	307
Db	499	RGAQIDENQADGDP-----YAFGRHQQKTTTGETPERFTY-----	538
QY	308	NLNGRESIINPGTAMASHKDEDEKFFPMSGVMIFGKESAGASANTALD-NVMTDDEEIIKA	366
Db	539	-----IAHQDT-----GRYPAGDMIQININFNLPVTNDNVLLP	570
QY	367	TNPVATERFGTVAVNFQSSSTDPAATGDVHAMGALPGWVQDRDYYLQGPIMAKIPIHTDGH	426
Db	571	TDPIG---GKTGINY--TNIFNTYGPLTALNNVP-----PVPNGQIWDKEFDTD--	615
QY	427	FHPSPLMGGFGLK-----NPPQILLIKNTPVPAN--PRAEFSAKFAFSIT	470
Db	616	-----LKPRLHVNAAPFVCQNNCPGQLFVKVAPNL'TNEYPDASANMSR-----IV	660
QY	471	QYSTQGVSEIEMELQKENSKRWNPEVQYTSNYAKSANDVFTVDNNGLYTEPRDIG	526
Db	661	TYSDFMWKGKLVFKAKLRASHTWNPICQMSIN-----VDNQFNYL-PANNIG	705

## RESULT 13

ВСПВІ





A:Residues: 1-727 <MAR>  
 A:Cross-references: GB:X55115  
 C:Superfamily: parvovirus coat protein  
 C:Keywords: coat protein  
 F:144-727/Product: coat protein VP2 #status predicted <VP2>

Query Match 6.5%; Score 189; DB 1; Length 727;  
 Best local Similarity 20.6%; Pred. No. 6.3e-06;  
 Matches 135; Conservative 73; Mismatches 216; Indels 232; Gaps 32;

```

QY 5 GGAPMADNNEGADGVGNASGNMHCSTWLGDRVITTTST-----RTWALPTYYN- 52
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 153 GGQP-AVRNERATGSGNGSGGGGGS--GGVGISTGTENNQTEFKPLENGWVEITANSS 209
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 53 ---HL-----YKQI-----SSASTGASNDNHVFGYSTPWGYPDFNRHCHFSPRDW 95
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 210 RLVLHLMPESENYKRVVNNMDKTAVKGNMALLDDIHVQIVTPWSLVDANAWGVWFNPGDW 269
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 96 QRLINNNWGFPRPKLNFKLFINIQVEVT---TNDGVTIANLTSVQVFSDEYQLPYV 152
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 270 QLVNTMSELHLVSFEQEIFNVVLKTVSESATQPTKVYNNDLTFASLMVALDSNNTMPFT 329
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 153 LGSAHQGL-----PPPA-----DVFMIPQGYLTINSGQAVGRS----- 189
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 330 PAAMRSETLGFPYWKPTIPTPRVYFQWDRLLPSH-----TGTSGPTNVVHGTDPDD 383
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 190 -SFYCLE-YFPSQMLRTGNF---TF-----SYTFEE-----VPHSSYAHQSOL 229
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 384 VQFYTIENSVPVHLIRTGDEFATGTFPFDCRCRLTHWTQTNRALGLPPLNSLPQSEGA 443
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 230 DR-----LMNPLIDQYL--YYLNRTQNS----- 251
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 444 TNFGDIGVQDDKRGVTOGMNTDYITEATINRPAEVGYSAPIYSFEASTQGPPEKPIAAG 503
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 252 -GSAQNKDLFSRGSFAGMSVQPKNWLPGPCY--RQQRVSKTKTDNNSNFTWTGASKY 307
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 504 RGAQTIDENQADGDP-----YAFGRQHGQKTTTGTETPERFTY----- 543
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 308 NLNGRESINPGTAMASHKDEDEKFFPMGVMIFGKESAGASNTALD-NVMITDEEERIKA 366
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 544 -----IAHQDT-----GRYPEGDWIQNINENLPTNDNVLLP 575
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 367 TNPVATERFGTVAVNFQSSSTDPAIGVHAMGALPGMWWQDRDVLVLOGPIWAKIPTHG 426
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 576 TDPIG---GKTGINY--TNIFNTYGPLTALNNVF-----PVYPNGQIWDKEPDTD-- 620
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 427 FHPSPIMGFGGLK-----NPPQILIKNTVPVPAV---PRAEFSATKFASTFT 470
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 621 -----LKPRLHVNAFPVCQNNCPGQLFVKVAPNLTNNEYDPDASANMSR-----IV 665
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 471 QYSTGVSVIEIWELOKENSKRWNPEVOYTSNYAKSANVDFVDNNGLYTEPRPIG 526
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 666 TYSDFWWMGKLVFKAKLRASHWTNPIQMSIN-----VDNQFNIV-PNNIG 710
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  
```

Search completed: July 17, 2003, 18:35:48  
 Job time : 21.1429 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2003, 09:15:11 ; Search time 7011 Seconds  
(without alignments)  
10898.630 Million cell updates/sec

Title: US-09-807-802A-1

Perfect score: 4718

Sequence: 1 ttgcgccactccctctctgcg.....cgcagagaggagtgggca 4718

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: em\_estbda:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	91.8	1.9	489	17	AZ289042 RPCI-23-5
2	76.2	1.6	753	17	BH115587 RPCI-24-3
3	71	1.5	495	17	AZ028362 RPCI-23-3
4	55.6	1.2	636	13	BM440166 pgrin.Dko
5	55.2	1.2	659	13	BM371945 EBma08_SQ
6	55.2	1.2	692	14	BQ766288 EBrc08_SQ

7	55.2	1.2	852	12	BG310300	BG310300 HVSMEC001
8	54	1.1	652	13	BJ268130	BJ268130
9	53.2	1.1	607	14	BQ660503	BQ660503 HI02005u
10	52.4	1.1	528	9	AL818251	AL818251 AL818251
11	52.4	1.1	688	14	BM816637	BM816637 HB02E06 T
12	52.4	1.1	1065	11	AY110092	AY110092 Zea mays
13	52	1.1	877	13	BI952167	BI952167 HVSME000
14	51.6	1.1	570	10	BE452865	BE452865 894066D03
15	51.6	1.1	612	13	BI873856	BI873856 963115H10
16	51.6	1.1	655	13	BI875610	BI875610 963125H10
17	51.6	1.1	680	12	BG844693	BG844693 1024007D1
18	51	1.1	536	10	BE603252	BE603252 HVSMEH010
19	50.6	1.1	500	12	BG905125	BG905125 TALR1137D
20	49.4	1.0	619	14	BQ822998	BQ822998 1030105D0
21	49.4	1.0	665	13	BI996796	BI996796 1031041H1
22	49.4	1.0	787	11	AY104224	AY104224 Zea mays
23	49	1.0	485	12	BF728378	BF728378 1000061E1
24	49	1.0	637	10	AM399968	AM399968 707055F01
25	49	1.0	686	13	BM610226	BM610226 170006871
26	49	1.0	874	11	AY104654	AY104654 Zea mays
27	48.8	1.0	599	12	BE726188	BE726188 894089F04
28	48.8	1.0	672	12	BG859092	BG859092 1024061C1
29	48.8	1.0	1137	13	BI952212	BI952212 HVSME000
30	48.6	1.0	495	12	BF630386	BF630386 HVSMEB000
31	48.2	1.0	592	12	BE776688	BE776688 MY-19-B-0
32	48	1.0	367	13	BJ432209	BJ432209 BU432209
33	48	1.0	559	9	AU093196	AU093196 AU093196
34	48	1.0	680	12	BE620370	BE620370 HVSMEC001
35	48	1.0	818	10	BE040681	BE040681 OF04H09 O
36	47.8	1.0	674	9	AI820329	AI820329 605091F03
37	47.6	1.0	621	17	BH868119	BH868119 hK22f04.Y
38	47.4	1.0	647	13	BM654776	BM654776 170006873
39	47.2	1.0	471	10	BE356945	BE356945 DG1_145_F
40	47.2	1.0	478	10	BE583083	BE583083 3-10F-MY
41	47.2	1.0	484	14	BM834231	BM834231 K-EST0109
42	47.2	1.0	576	14	BM846866	BM846866 K-EST0125
43	47.2	1.0	592	10	BE584361	BE584361 10-1D-HA
44	47.2	1.0	597	13	BI818181	BI818181 603032640
45	47.2	1.0	609	10	BE583082	BE583082 3-9A-MY P

ALIGNMENTS

RESULT 1  
AZ289042 489 bp DNA linear GSS 27-JUL-2000  
LOCUS RPCI-23-59A6.TVB RPCI-23 Mus musculus genomic clone RPCI-23-59A6,  
DEFINITION DNA sequence.  
ACCESSION AZ289042  
VERSION AZ289042.1 GI:9530917  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 489)  
AUTHORS Zhao,S., Niernan,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret,  
'B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.  
and Fraser,C.M.  
TITLE Mouse BAC End Sequences from Library RPCI-23  
JOURNAL Unpublished (1999)  
COMMENT Other\_GSSs: RPCI-23-59A6.TVB  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org  
Clones are derived from the mouse BAC library RPCI-23. For BAC  
library availability, please contact Pieter de Jong  
(pieter@dejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>)  
or from Resea ch Genetics ([info@resgen.com](mailto:info@resgen.com)). BAC end page:  
[http://www.tigr.org/tldb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html)  
plate: 59' row: A column: 6  
Seq primer: T7  
Class: BAC ends.

FEATURES  
source  
location/Qualifiers

1.489  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone\_1ib="RPCI-23-59A6"  
/sex="Female"  
/lab\_host="DH10B"  
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site\_1:  
ECORI; Site\_2: EcorI; Female C57BL/6J mouse kidney and/or  
brain genomic DNA was isolated and partially digested  
with a combination of EcorI and EcorI Methylase. Size  
selected DNA was cloned into the pBACe3.6 vector at the  
EcorI sites. The ligation products were transformed into  
DH10B electrocompetent cells (BRL Life Technologies)."  
BASE COUNT 138 a 111 c 104 g 136 t  
ORIGIN

Query Match 1.9%; Score 91.8; DB 17; Length 489;  
Best Local Similarity 56.5%; Pred. No. 6.5e-12;  
Matches 264; Conservative 0; Mismatches 177; Indels 26; Gaps 4;

OY 1332 TTGGCGCGCCACCACGCGGCAAGACCAATCGCGGAAGCCATCGCCACGCCGTGCCCT 1391  
DB 13 TTAGCCTGGCTACTATGAGAAACAAGTGTTCAGTATCTTACACATGCTGTTTCAT 72  
OY 1392 TCTACGCGTGGCTCACTGAGCAATGAGAACTTCCCTTCAATGATGCGT-----CG 1445  
DB 73 TCTATGATTTGTGATCTGGATCATATAAGACATTCCTTCAATGCGCTGCCAGTTTCCCA 132  
OY 1446 ACAAGATGCTGATCTGTGGAGGAGGAGGCAAGATGACGGCCCAAGTCTGTGAGTCCGCCA 1505  
DB 133 TCAAAATGCTCATTTGTGTAATAAGAGGGAAGATGTGTCCAAGATGTTCAATTCTGCCA 192  
OY 1506 AGGCCATTCTCGCGCGGACGAAGGTGCGCGTGGACCAAAAGTGCAGATCGTCCGCCAGGA 1565  
DB 193 AGAGCATTTTCACTGCTTGA-----GTTGTCAAAAAGTGCAGTACTTGAATGGA 244  
OY 1566 TCGACCCACCCCGCGTATGTCACCTCCACCAACATGTGCGCGGTGATGACGGGA 1625  
DB 245 TCAACAGCACTCAGTCAAGCATCACCAGAACACTGACATGACCCTTGCCCTCAACTTCA 304  
OY 1626 ACAGACCACTTTCAGACACCAAGCCGTTGAGGACCGGATGTTCAATTGAACTCA 1685  
DB 305 ATGTCA-----CAAGCATCACCAAGCCCTTGAAGATAGATGT-----TCTACCTTG 352  
OY 1686 CCGCGCTGTGAGATGACTTTGGCAAGGTGACAAAGCAGAGTCAAGATTCTTCC 1745  
DB 353 GCAGAGACTCCAGCAAGACCTTGGCAGATTTCAATGGAGTGCACCAAGTTCTTTA 412  
OY 1746 GCTGGCGCAGATCACTGACCGAGGTGGCGCATGAGTTCTACGTC 1792  
DB 413 AATGATGATCGTAGTGAATTCTGTGTGACCAATATGCTCTGCTG 459

RESULT 2  
LOCUS BH115587 753 bp DNA linear GSS 19-JUL-2001  
DEFINITION RPCI-24-358F16.TV RPCI-24 Mus musculus genomic clone RPCI-24-358F16  
ACCESSION BH115587  
VERSION BH115587.1 GI:14954954  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 753)  
Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,  
Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,  
Russell,D., de Jong,P. and Fraser,C.M.  
Mouse BAC End Sequences from Library RPCI-24  
Unpublished (1999)  
Other GSSs: RPCI-24-358F16.TJ  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org  
Clones are derived from the mouse BAC library RPCI-24. For BAC  
library availability, please contact Pieter de Jong  
(pdejong@mail.cho.org). Clones may be purchased from BACPAC  
Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end  
page: [http://www.tigr.org/tldb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html)  
plate: 358 row: F column: 16  
Seq primer: T7  
Class: BAC ends.

FEATURES  
source  
location/Qualifiers

1.753  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="RPCI-24-358F16"  
/clone\_1ib="RPCI-24"  
/sex="Male"  
/cell\_type="Spleen/Brain"  
/note="Vector: pTARBAC1; Site\_1: BamHI; Site\_2: BamHI;  
RPCI-24 Mouse BAC library produced by Pieter de Jong. The  
library was cloned in the pTARBAC1 cloning vector at the  
BamHI sites using MboI partially digested male C57BL/6J  
DNA."  
BASE COUNT 211 a 178 c 167 g 197 t  
ORIGIN

Query Match 1.6%; Score 76.2; DB 17; Length 753;  
Best Local Similarity 51.6%; Pred. No. 6.2e-08;  
Matches 268; Conservative 0; Mismatches 223; Indels 28; Gaps 3;

OY 2488 CGTACGCGGTATTAACACGCGGACGCCGAGTTTCAAGAGCGTCTGCAAGAAGATACGT 2547  
DB 207 CTGTCTCCATAAATACACAGACAGAGTTTCAAGAGAAACTCCAGTACAAATAC 266  
OY 2548 CTTTGGGGGCAACCTCGGGCGACAGTCTTCCAGGCCAAGAACGGGTCTCGAACCTC 2607  
DB 267 TTTTTTTGGCA-----GGAACCGTGCCCAATGCGCAAGAAAGGGTCTCAAAACCA 318  
OY 2608 TCGGTCTGTGAGGAAGCGCTTAAGACGGCTCTCTGGAAGAAACGTCGGTAGAGCAGT 2667  
DB 319 TCGGCTTGCTTAAGAGGATTCAGCATGGCTCTCGGAAGAAAT----- 363  
OY 2668 CGCCACAAGACCAAGACTCTCTCGGGCATCGGCAAGACAGCCAGCGCGTAAAA 2727  
DB 364 ----AGATTGCACTTAAACTCTCCCCACATGAAGAGTACACTCGCTCTTTTCAGA 419  
OY 2728 AGAGACTCAATTTTGTGAGACTGGGACTCAGAGTCACTCCCGATCCACCAACCTCTCG 2787  
DB 420 GTTATGCCAAAAACAGTAGAGTGAAGAGCTGGCTGCTATAGTCTATGACAACAGATGTA 479  
OY 2788 GAGAACTCCAGCAACCCCGCTGCTGTGGAGCTACTATCAATGCTTCAAGCGGTGGCG 2847  
DB 480 GCAGTTTCCAGCAGATATCTTCAATTAGACACTTCTATAATCTCTGAGCTGAGGTC 539  
OY 2848 CACCAATGGCAGACAATAACGAAGCGCCGACGAGAGTGGGTAATGCCCTCAGGAATGGC 2907  
DB 540 TACTTAATGGGCAACAACAAC-CAGACACTGATGAGTGAAGCAATGCCATGATGATGGC 598  
OY 2908 ATTGCATTCACATGCTGGGGGACAGAGTCAATCACCACCAAGACCCGACCTGGGCT 2967



Db 599 ATTGCCATTCAGTGTATGTAGACTGATGTCATTACCTAATCTCCAGACCTGGTCC 658  
QY 2968 TGCCCACTTACATAAACCACCTCTACAGCAATCTCCA 3006  
Db 659 TGCCCACTTCAACTAATTCACACTATATGACA 697

## RESULT 3

AZ028362/c 495 bp DNA linear GSS 25-FEB-2000  
LOCUS RPCI-23-317G13.TJ RPCI-23 Mus musculus genomic clone RPCI-23-317G13  
DEFINITION ' DNA sequence.  
ACCESSION AZ028362  
VERSION AZ028362.1 GI:7103746  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
AUTHORS 1 (bases 1 to 495)  
Zhao, S., Nierman, W., Feldblum, T., Malek, J., Shatsman, S., Akinret  
, B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.  
and Fraser, C. M.

TITLE Mouse BAC End Sequences from Library RPCI-23  
JOURNAL Unpublished (1999)  
COMMENT Other GSSs: RPCI-23-317G13.TV  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC  
library availability, please contact Pieter de Jong  
(pieter@dejong.med.buffalo.edu). Clones may be purchased from  
BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>)  
or from Resea ch Genetics ([info@resgen.com](mailto:info@resgen.com)). BAC end page:  
[http://www.tigr.org/tdb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html)  
Plate: 317 row: G column: 13  
Seq primer: SP6  
Class: BAC ends.

FEATURES  
source Location/Qualifiers

1. 495  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="RPCI-23-317G13"  
/clone\_lib="RPCI-23"  
/sex="Female"  
/lab\_host="DH10B"

/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site\_1:  
EcoRI; Site\_2: EcoRI; Female C57BL/6J mouse kidney and/or  
brain genomic DNA was isolated and partially digested  
with a combination of EcoRI and EcoRI Methylase. Size  
selected DNA was cloned into the pBACe3.6 vector at the  
EcoRI sites. The ligation products were transformed into  
DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 124 a 109 c 92 g 170 t

ORIGIN

Query Match 1.5%; Score 71; DB 17; Length 495;  
Best Local Similarity 56.1%; Pred. No. 1.1e-06;  
Matches 134; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

QY 1031 GACCGGGGATCACTCCGAGAGCAGTGCAGAGGAGCAGGCTCGTACATCTCC 1090  
Db 242 GAAGAGGTAATTTCTATCAGAGAGGAATGGTCCATAAAAACCAAGATCAGAACCATCT 183  
QY 1091 TTCAACGCCGCTTCCAACTCGCGGTCCAGATCAAGGCCGCTTGGAATGCCGCAAG 1150  
Db 182 TACAATATCGCAGTTCCAACCAAGATCAATCAAAAGTGCTTGCAATGTCAACCAAG 123

QY 1151 ATCATGGCGCTGACCAATCCGCGCCGACTACCTGTAGGCCCCGCTCCGCGCGAC 1210  
Db 122 ATTGTGACTGTGACCAAGTAGATGGAAGATTAGCTGATTGGCAAGATCCACTGGCAGAT 63  
QY 1211 ATTAAACCAACCGCATCTACCGCATCTCTGGAGCTGAACGGCTACGAACCTGCCTACGC 1269  
Db 62 GTCTCCTCCAGTCACATTACAGATATTTTGATGATGCTATGATCCAGTTTAGC 4

## RESULT 4

BM440166 636 bp mRNA linear EST 01-FEB-2002  
LOCUS BM440166  
DEFINITION pgrln.pk002.f22 Normalized Chicken Reproductive Tract cDNA library  
(pgrln) Gallus gallus cDNA clone pgrln.pk002.f22 5' similar to  
gb|AAH07597.1|AAH07597 (BC007597) Similar to plectin 1,  
intermediate filament binding protein, 500kd [Homo sapiens], mRNA  
sequence.

ACCESSION BM440166 GI:18470941  
VERSION BM440166  
KEYWORDS EST.  
SOURCE chicken.  
ORGANISM Gallus gallus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.  
1 (bases 1 to 636)  
Cogburn, L. A. and Nys, Y.

TITLE ESTs from Normalized Chicken Reproductive Tract cDNA library-  
University of Delaware and INRA, Tours-Poultry Unit Project  
JOURNAL Unpublished (2002)  
COMMENT Contact: Larry A. Cogburn  
University of Delaware  
Townsend Hall, Newark, DE 19717, USA  
Tel: 302-831-1335  
Fax: 302-831-2822  
Email: [cogburn@udel.edu](mailto:cogburn@udel.edu), [www.chickest.udel.edu](http://www.chickest.udel.edu).

FEATURES  
source Location/Qualifiers

1. 636  
/organism="Gallus gallus"  
/strain="Commercial broiler and layer"  
/db\_xref="taxon:9031"  
/clone="pgrln.pk002.f22"  
/clone\_lib="Normalized Chicken Reproductive Tract cDNA  
library (pgrln)"  
/sex="Male and Female"  
/tissue\_type="Testis, ovary and oviduct"  
/dev\_stage="Various stages; embryonic, post-hatch, immature  
and sexually-mature"  
/lab\_host="E. coli EMD10B"

/note="Vector: pCMVSPORT6; Library made from three total  
RNA pools from each tissue (testis 25%, ovary 25%, and  
oviduct 50% of final RNA pool); Single pass sequencing  
from 5'-end"

BASE COUNT 103 a 216 c 220 g 89 t 8 others

ORIGIN

Query Match 1.2%; Score 55.6; DB 13; Length 636;  
Best Local Similarity 46.1%; Pred. No. 0.0087;  
Matches 224; Conservative 0; Mismatches 259; Indels 3; Gaps 1;

QY 1218 CCAACCGCATCTACCGCATCTGAGCTGAACGGCTACGAACCTGCTACGCGGCTCG 1277  
Db 24 CTTGCCACCGCGCTTGTCTACACAAGAGACGGCGACTGTGCGGAGCCAGCGAGG 83  
QY 1278 TCTTTCTCGGCTGGGCCCAAAAGTTGGAAGCGCAACCATCTGCTGTTGGGC 1337  
Db 84 TGGCAGCTTCCCGGACCGCTCCACCGAGAGAGGCTGACTACACGAGCTGCTGGGC 143  
QY 1338 CGGCCACCAAGGGAAGACCAACATCGCGGAAGCCATCGCCACGCGCTTCTACG 1397  
Db 144 GCTGCCCAAGAGACGAGGGGACGGGCTGCTGCTGCTGCTGGGCGAGCGCGCGCC 203







ACCESSION BJ268130  
VERSION BJ268130.1 GI:20088244  
KEYWORDS EST.  
SOURCE bread wheat.  
ORGANISM Triticum aestivum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae ; Triticeae; Triticum.  
1 (bases 1 to 652)  
Ogihara, Y. and Murai, K.  
Expressed genes in Triticum aestivum  
Unpublished (2002)  
JOURNAL  
COMMENT Contact: Tadasu Shin-i  
Center For Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshini@genes.nig.ac.jp.

FEATURES  
source  
1. .652  
/organism="Triticum aestivum"  
/cultivar="Chinese Spring"  
/db\_xref="taxon:4565"  
/clone="whoh15f01"  
/clone\_lib="Y. Ogihara unpublished cDNA library, wh\_oh"  
/tissue\_type="pistil at heading date"  
/dev\_stage="Peekes' scale 10.5"  
/note="Vector: Lambda Uni-ZAP XR, excised phagemid; Site 1: EcoRI; Site 2: XhoI; Plants were grown under hydroponic conditions at UC Davis, salt stressed for 12 hours, and for 7 days, then dissected and frozen (Akhunov in J Dvorak Lab). Total RNA was prepared from sheath tissue, equal quantities of RNA were pooled from the two samples, polyA was purified from the pooled RNA, a cDNA library was made, and the cDNA clones were in vivo excised to give Bluescript phagemids in the TU Close lab at the University of California, Riverside (Akhunov, Chin , Choi, Close, Fenton, Kianian, Otto, Simons, Zhang). Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."

BASE COUNT 166 a 209 c 189 g 85 t 3 others  
ORIGIN

Query Match 1.1%; Score 54; DB 13; Length 652;  
Best Local Similarity 49.3%; Pred. No. 0.022;  
Matches 141; Conservative 0; Mismatches 145; Indels 0; Gaps 0;

QY 1384 CGTGCCCTTCTACGGCTGCGTCACTGACCAATGAACTTCCCTTCATGATTGCGT 1443  
DB 237 CATACACTCGACGGCTGCAAGCACAAGGTCAAGAGCTGCTCCAGAGATCGAAGCGT 296  
QY 1444 CGACAAGATGTTCTGTTGGAGAGGGGCAAGATGACGGCCAAAGTCTGAGTCCGC 1503  
DB 297 CTACTCGGTGGCCATCGATGTGCAACACCAAGTCAAGGTGACCGGCACGCTGACTC 356  
QY 1504 CAAGCCATTCTCGCGGGCAGCAAGGTGCGGTGGAACCAANGTCAAGTGTCCGCCA 1563  
DB 357 TGAGACGTGATCGGAAGCTCACCAGAGGGGCAAGCAGCAGAGCTGTGTGACCA 416  
QY 1564 GATGACCCACCCCGGTGATGTCACCTCCAACAACATGTGCGCGTGAATTGACGG 1623  
DB 417 GAAGGGCGCAACGGCGGCAACAACAACCAAGGGCCACANGGCGGCGGCGGCGG 476  
QY 1624 GAACAGACCACTTCGAGCAGCAGCAGCGGTGAGACCGGATG 1669  
DB 477 CAACAACAACGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 522

RESULT 9  
BQ660503 607 bp mRNA linear EST 15-JUL-2002  
LOCUS BQ660503  
DEFINITION HI02005u HI Hordeum vulgare cDNA clone HI02005 3-PRIME, mRNA

sequence.  
ACCESSION BQ660503  
VERSION BQ660503.1 GI:21801636  
KEYWORDS EST.  
SOURCE Hordeum vulgare.  
ORGANISM Hordeum vulgare  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae ; Triticeae; Hordeum.  
1 (bases 1 to 607)  
Michalek, W., Weschke, W., Pleisner, K.-P. and Graner, A.  
EST sequencing and analysis in barley  
Unpublished (2000)  
JOURNAL  
COMMENT Contact: Stein Nils  
Molecular Markers Group, Department Genbank  
Institute of Plant Genetics and Crop Plant Research (IPK)  
Corrensstr. 3, 06466, Gatersleben, Germany  
Tel: 039482-5522  
Fax: 039482-5595  
Email: stein@ipk-gatersleben.de  
Insert Length: 607 Std Error: 0.00  
Plate: 2 row: 0 column: 5  
Seq primer: M13uni.

FEATURES  
source  
1. .607  
/organism="Hordeum vulgare"  
/cultivar="Barke"  
/db\_xref="GABI:136895"  
/db\_xref="taxon:4513"  
/clone="HI02005"  
/clone\_lib="HI"  
/tissue\_type="female inflorescences"  
/dev\_stage="female inflorescences (approx. 3 mm in size)"  
/lab\_host="XL10-Gold"  
/note="Vector: Bluescript SK+; Site 1: EcoRI (5'-end of cDNA); Site 2: XhoI (3'-end of cDNA); Due to a cloning artefact caused by the kit, in most cases the EcoRI site is NOT present, as well as the EcoRI adapter used for cloning. To excise the insert, restriction sites upstream EcoRI should be used (e.g. BamHI, SalI, PstI). NOTE: Also due to the cloning system used Blue/white selection for recombinants is not 100% reliable."

BASE COUNT 132 a 251 c 162 g 62 t  
ORIGIN

Query Match 1.1%; Score 53.2; DB 14; Length 607;  
Best Local Similarity 47.9%; Pred. No. 0.034;  
Matches 185; Conservative 0; Mismatches 198; Indels 3; Gaps 1;

QY 1078 CTGCTACATCTCTCTCAACGCGGCTTCCAACTCGCGGTCCAGATCAAGCGCTCTGA 1137  
DB 205 CGCCGACCGCGGCTCTTCTCCCTCCCGCGCGCCGCAACAACGCGCTTCCAA 264  
QY 1138 CAATGCCGCAAGATCATGGCGCTGACCAAAATCCGCGCCGACTAAGCTG--TAGGCC 1194  
DB 265 GCCCAACAACAGCTCAACGCGCGGCTCATCGGTCGGCAAGCTCGCCTCGGCGC 324  
QY 1195 CGCTCGCGCGGACATTAAACCAACCGCATCTACCGCATCTGAGCTGAACGCTA 1254  
DB 325 CACCACCAACGAGGCGGACCGCTACACAACGTCAGCTCCCGCAAGCCCTA 384  
QY 1255 CGAAGCTGCTACGCGGCTCGCTTTCTCGGCTGGGCCAGAAAGTTGGAAGCG 1314  
DB 385 CGCCACGCAACGCGGCTGTAAGGCGGCAAGCCCAACGCAACATCAACGG 444  
QY 1315 CAACACCATCTGCTGTTGGGCGGCAACGAGGCAAGCAACATCGCGGAAGCAT 1374  
DB 445 CAACACTTGGGGCTCAGCAAGATGGCGGCAAGCGGCGGCGGCTACAGCAACTCAA 504  
QY 1375 CGCCACGCGGCTGCTCTACGGCTGCTCACTGAGCAATGAGAACTTCCCTCAA 1434  
DB 505 CGCGGCAAGAGGGGTCAAGAGCTACTTCAACAAGTCCATCGGAGCGCGGCAAGCA 564

Qy 1435 TGATTGCGTCGACAAGATGTGATCT 1460  
Db 565 CAACAGCAACAGCAGCACTTCAACT 590

RESULT 10

LOCUS AL818251 528 bp mRNA linear EST 15-JUL-2002  
DEFINITION AL818251 j:223 Triticum aestivum cDNA clone C05\_j223\_plate\_14, mRNA sequence.

ACCESSION AL818251  
VERSION AL818251.1 GI:21829851  
KEYWORDS EST.  
SOURCE bread wheat.  
ORGANISM Triticum aestivum

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae ; Triticeae; Triticum.  
1 (bases 1 to 528)

AUTHORS Wilson, I., Beswick, R., Shepherd, S., Barker, G., Parker, J., Owen, P., Edwards, D., Coghill, J., Holdsworth, M., Lenton, J., Shewry, P. and Edwards, K.

TITLE A BBSRC-funded wheat EST resource for the academic community  
JOURNAL Unpublished (2002)  
COMMENT Contact: Barker G  
Institute of Arable Crop Research  
Long Ashton, Bristol BS41 9AF United Kingdom.

FEATURES  
source 1. 528  
/organism="Triticum aestivum"  
/cultivar="mercia"  
/db\_xref="taxon:4565"  
/clone="C05\_j223\_plate\_14"  
/clone\_lib="j:223"  
/tissue\_type="embryo"  
/dev\_stage="21 days post anthesis"  
location/Qualifiers

BASE COUNT 138 a 169 c 157 g 64 t  
ORIGIN

Query Match 1.1%; Score 52.4; DB 9; Length 528;  
Best Local Similarity 49.0%; Pred. No. 0.052;  
Matches 140; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

Qy 1384 CGTGCCTTCTACGGCTGCGTCACTGACCAATGAGACTTCCCTTCATGATTGCGT 1443  
Db 195 CATACACTGCGACGGCTGCAAGCACAAGGTCAAGAAGCTGCTCAGAGATCGAAGCGT 254  
Qy 1444 CGACAAGATGTGATCTGTGGAGGAGGCAAGATGACGGCCAAAGTCTGTGAGTCCGC 1503  
Db 255 CTACTCGGTGGCCATGATGTGACACAACAAGTCACGGTGACCGGACGCTGACTC 314  
Qy 1504 CAAGGCCATTCTCGGGCGCAGCAAGGTGGCGGTGACCAAAAGTGCAAGTCTCCGCCA 1563  
Db 315 TGAGACGCTGATCCGGAAGCTCACAGAGGGGGCAAGCAGCAGCTGTGTCCACCA 374  
Qy 1564 GATGACCCCAACCCCGGTGATGCTCACTCCCAACCAACATGTGCGCCGTGATTGACGG 1623  
Db 375 GAAGGGCGCAACGGCGGCAACAACATACTACAGGGCCACAAGGGCGGGCGGGCGG 434  
Qy 1624 GAACAGACCACTTTCGAGCAGCAGCAGCCGTTCAGAGACCGGATG 1669  
Db 435 CAACCAACAACGACGAGAGCCGCGCAGCAGCAGCAGCAGCAGAG 480

RESULT 11

BM816637

LOCUS

DEFINITION BM816637 688 bp mRNA linear EST 05-MAR-2002  
HB02E06 T3.ab1 HB Hordeum vulgare cDNA clone HB02E06 T3.ab1 similar to Arginine decarboxylase, Arginine 2-monooxygenase, Glycine dehydrogenase (decarboxylating), mRNA sequence.

ACCESSION

VERSION BM816637

KEYWORDS BM816637.1 GI:19152651 EST.

SOURCE Hordeum vulgare.  
ORGANISM Hordeum vulgare

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae ; Triticeae; Hordeum.  
1 (bases 1 to 688)

AUTHORS Ozurk, N.Z., Michalowski, C.B., Brazille, S., Borchert, C., Palacio, C., Normand, C., Murphy, C., Kelley, R., Sant, S.A., McLaughlin, H., Fredricksen, M.A. and Bohnert, H.J.

TITLE Monitoring large-scale changes in transcript abundance in drought- and salt-stressed barley

JOURNAL Unpublished (2002)  
COMMENT Contact: Mark A. Fredricksen  
Plant Biology

University of Illinois  
1201 W Gregory Dr, Urbana, IL 61801, USA  
Tel: 2172655473

Email: bohnertlab@life.uiuc.edu.  
location/Qualifiers

FEATURES  
source 1. 688

/organism="Hordeum vulgare"  
/strain="cv tokak"  
/db\_xref="taxon:4513"  
/clone="HB02E06\_T3.ab1"  
/clone\_lib="HB"  
/tissue\_type="leaf"  
/dev\_stage="3 week old"  
/note="6 and 10 hour drought stress by placing plants on moist paper (75% rel. humidity) in light"

BASE COUNT 181 a 258 c 157 g 89 t 3 others  
ORIGIN

Query Match 1.1%; Score 52.4; DB 14; Length 688;  
Best Local Similarity 48.1%; Pred. No. 0.057;  
Matches 180; Conservative 0; Mismatches 191; Indels 3; Gaps 1;

Qy 1078 CTCGTACATCTCTTCAACGCGCTTCCAACTCGCGGTCCAGATCAAGCGCTCTGA 1137  
Db 153 CGCCGACCCCGGCTCTTCTCCCTCCCGCCGCGCCCAACAACGCGCTTCCA 212  
Qy 1138 CAATGCGGCAAGATGCGCGCTGACCAATCCGCGCCGACTACTG--TAGGCC 1194  
Db 213 GCCCAACAACAGCCTCAAGGCCCGGCTCATCGGTCCGCAAGCTCGCTCGGCC 272  
Qy 1195 CGCTCCGCGCGGACATTAACAACCGCATCTACCGCATCTGAGCTGAACGCTA 1254  
Db 273 CACCAACCAACGCAAGGCCGACCGCTACACAACAGTCAAGCTCCCGACGCAAGCCTA 332  
Qy 1255 CGAAGCTGCTAGCGCGGCTCTTCTTCTCGGCTGGGCCAGAAAAGTTCCGGAAGCG 1314  
Db 333 CGCCAACGCAACGCGGCTAGCGGCTACGCGCCAAACCAACGCAACATCAACGG 392  
Qy 1315 CAACACCATCTGGCTGTTGGGCGGCCACCAAGGCAAGCAACATCGCGAAGCAT 1374  
Db 393 CAACACCTTGGGCTCAACAAGATGGCGGCAAGCGGCGGCGGCTACAGCAACTTCAA 452  
Qy 1375 CGCCCAAGCGGCTGCTTCTACGCGCTGCTCAACTGGAACCAATGAACTTCCCTTCAA 1434  
Db 453 CGGGGCAACGAGGGGGTCAAGAGTACTTCAACAGTCCATCGGACGCGCGGCA 512  
Qy 1435 TGATTGCGTCGACA 1448  
Db 513 CAACAGCAACAGCA 526

RESULT 12

AY110092/c

LOCUS AY110092 1065 bp mRNA linear HTC 25-MAY-2002  
DEFINITION Zea mays CL49602\_1 mRNA sequence.

ACCESSION AY110092

VERSION AY110092.1 GI:21214182

KEYWORDS HTC.  
Zea mays.

ORGANISM      Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
clade; Panicoideae; Andropogoneae; Zea.  
REFERENCE  
AUTHORS      Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitesitt,M.S.,  
Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.  
TITLE          Maize Mapping Project/DuPont Consensus Sequences for Design of  
Overgo Probes  
JOURNAL        Unpublished (2002)  
REFERENCE      2 (bases 1 to 1065)  
AUTHORS        Coe,E.C.  
TITLE          Direct Submission  
JOURNAL        Submitted (25-APR-2002) Maize Mapping Project, University of  
Missouri, Columbia, MO 65211, USA  
FEATURES  
source  
1.1065  
/organism="Zea mays"  
/db\_xref="MaizeDB:632202"  
/db\_xref="taxon:4577"  
/clone="CL49602.1"  
/clone\_lib="Maize Mapping Project/DuPont Consensus  
Library"  
/note="this sequence is part of a project of EST  
assemblies resulting from the application of public  
contigs to seed DuPont contigs; this resource was  
assembled by DuPont as part of a collaboration for the  
overgo addressing of BACS in conjunction with the Maize  
Mapping Project"

BASE COUNT      178 a      300 c      260 g      221 t      106 others  
ORIGIN

Query Match      1.1%; Score 52.4; DB 11; Length 1065;  
Best Local Similarity 43.7%; Pred. No. 0.065;  
Matches 204; Conservative 0; Mismatches 260; Indels 3; Gaps 1;

QY 1308 GGAAGCGCAACACCATCTGCGTGTGGCGCGCCACGAGGCGCAAGCAATCGCGG 1367  
| | | | |  
DB 916 GCAAGCGNN 857  
| | | | |  
QY 1368 AAGCCATCGCCACGCGCGTCTTCTACGCGTGGTCAACTGACCAATGAACTTTC 1427  
| | | | |  
DB 856 AGGAGCGCGATTCGTGCGCTCCTCGACGCGGAGTCAAGTTCAACGCTTCTTCC 797  
| | | | |  
QY 1428 CCTCATGATTCGTCGACAGAT--GGTATCTGTGGAGGAGGCAAGATGACGG 1484  
| | | | |  
DB 796 TCGAGAAGAGAGAGACTACGTATCCGCGTGAAGAGTTGACAGACAGGCGTGAGCG 737  
| | | | |  
QY 1485 CCAAGTCTGAGTCCGCCAAGGCCATTTCTGCGCGCAGCAAGTGCCTGGAACCAA 1544  
| | | | |  
DB 736 CCGCGGAGGTGGGGTGGCGGAGGAGCTGCTGCGGGTGGCAGGAGATCTGACTTCC 677  
| | | | |  
QY 1545 AGTGAAGTCTCCGCCAGATGACACCCCGGCGTATGTCACCTCCAACCAACA 1604  
| | | | |  
DB 676 ACGCGAGATGCTTGTCTGAGAACTACAGCGCGCTCAACTACACCGGCTGTCAAGA 617  
| | | | |  
QY 1605 TGTGCGCGTGAATGACGGAACAGACACCACTTCAGACACGAGCGCGTTGAGACC 1664  
| | | | |  
DB 616 TCCTCAAGAGTACGACAGAGACCGCGCGCTGATCCGCTTCAATCCGAACG 557  
| | | | |  
QY 1665 GGATGTTCAATTTGAATCAACCGCGCTGTGAGCATGACTTTGGCAAGGTGACAAAGC 1724  
| | | | |  
DB 556 TGATGACGAGCGCGTTCTGCGCCACGACGTGCTTACAAGTCTGTCAAGGGTGCGAGG 497  
| | | | |  
QY 1725 AGGAAGTCAAGAAGTTCTTCCGCTGGCGCGCAGAGTCACTGACCGAG 1771  
| | | | |  
DB 496 AGATGCTGACCAAGCTCTCTCTGCCCCGGAACCATCAAGCGGCCCGTG 450  
| | | | |

RESULT 13  
BI952167      877 bp      mRNA      linear      EST 19-OCT-2001  
LOCUS      BI952167  
DEFINITION      HVSMem0004N17f Hordeum vulgare green seedling EST library.

ACCESSION  
VERSION      HVCDNA0014 (Blumeria infected) Hordeum vulgare cDNA clone  
BI952167  
HVSMem0004N17f, mRNA sequence.  
KEYWORDS  
SOURCE        BI952167.1      GI:16295739  
ORGANISM  
REFERENCE  
AUTHORS        Hordeum vulgare.  
TITLE          Hordeum vulgare  
JOURNAL        Unpublished (2001)  
COMMENT        Contact: Wing RA  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: rwing@clemson.edu  
Total hg bases = 467  
Seq primer: AATTACCCCTCACTAAGGG  
High quality sequence stop: 849.  
FEATURES  
source  
1.877  
/organism="Hordeum vulgare"  
/cultivar="Morex"  
/db\_xref="taxon:4513"  
/clone="HVSMem0004N17f"  
/clone\_lib="Hordeum vulgare green seedling EST library  
HVCDNA0014 (Blumeria infected)"  
/tissue\_type="green seedling leaf"  
/lab\_host="TJ121"  
/note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:  
XhoI; Morex (mla) plants were greenhouse grown in the R  
Wise lab at Iowa State University, Ames, IA; 7 day old  
green seedlings were infected with isolate 5874 of  
Blumeria graminis f. sp. hordei, and leaves were harvested  
24, 48 and 72 hr post-inoculation and snap frozen (Wise).  
In the TJ Close lab at the University of California,  
Riverside, total RNA was prepared from each sample pool,  
equal quantities of all three RNA pools were combined,  
poly(A) RNA was purified from the mixture, one primary  
unamplified cDNA library was made, and 1 million pfu were  
in vivo excised to give pBluescript SK(-) cDNA phagemids  
(Chin). Phagemids were plated and picked at the Clemson  
University Genomics Institute (CUGI) (Begum, Palmer,  
Frisch, Atkins and wing). Plasmid DNA preparations, DNA  
sequencing and sequence analysis were performed at CUGI  
(Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main).  
The sequence has been trimmed to remove vector sequence  
and contains a minimum of 100 bases of phred value 20 or  
above. For more details on library preparation and  
sequence analysis see  
http://www.genome.clemson.edu/projects/barley. To order  
this clone see http://www.genome.clemson.edu/orders Also  
see Close TJ, Wing R, Kleinofs A, Wise R (2001)  
Genetically and physically anchored EST resources for  
barley genomics. Barley Genetics Newsletter 31:29-30.  
(http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"

BASE COUNT      192 a      319 c      259 g      105 t      2 others  
ORIGIN

Query Match      1.1%; Score 52; DB 13; Length 877;  
Best Local Similarity 47.1%; Pred. No. 0.077;  
Matches 194; Conservative 0; Mismatches 215; Indels 3; Gaps 1;

QY 1078 CTCGTACATCTCTTCAACGCGCGTCCAACTGCGGCTCCAGATCAAGCGCGCTTGGA 1137  
| | | | |





mid-log phase in TAP-N (30 min, 1hr, 4hr), TAP-S (30 min, 1hr, 4hr), TAP-P (4hr, 12hr, 24hr), NO3 to NH4 (30min, 1hr, 4hr) and NH4 to NO3 (30min, 1hr, 4hr). PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda Zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites. Bluescript II SK- plasmids were excised from the lambda Zap clones by superinfection with ExAssist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al (1996) Genome Research 6: 791-806."

BASE COUNT 144 a 190 c 180 g 98 t  
ORIGIN

Query Match 1.1%; Score 51.6; DB 13; Length 612;  
Best Local Similarity 46.0%; Pred. No. 0.087;  
Matches 174; Conservative 0; Mismatches 204; Indels 0; Gaps 0;

QY	1332	TTGGGCGCGCCACCAAGGCAAGACCAATCGCGGAAGCCATCGCCACGCGTGCCCT	1391
DB	191	TTGGCAAGAGGGCGAGATCAAGACCGTGCTGTGGGCTACTGGCGCACTTCTGCTGC	250
QY	1392	TCTACGGCTGCGTCAACTGGAACCAATGAGACTTCCCTTCAATGATGTCGTCACAAGA	1451
DB	251	CCAACGGCATGGCCAGATGTCAGAGGGGATCTCAACCAATCCGCGCAAGAGG	310
QY	1452	TGATGATCTGTGGAGAGAGGCAAGATGACGCGCAAGTCGTGAGTCCGCCAAGGCCA	1511
DB	311	AGGACGATCCGCAAGAGCTGAGAGAGAGGCCAGGCGAGGCTTTGCCAAGCCCC	370
QY	1512	TTCTGCGCGCAGCAGGTGCGCGTGACCAAAAGTCAAGTCGTCCGCCAGATCGACC	1571
DB	371	TGACCACCATCGGCAAGTCTCTGTAAGAAGAGACCGCGAGAGACCAAGATCTACG	430
QY	1572	CCACCCCGTATGCTCACCTCCACACCAACATGTGCGCGTGATGACGGGACAGCA	1631
DB	431	GCTCCGTGACGGTGACAGAGATGCGGACGCCATCTACAGACAGACCGCGCAACGTGA	490
QY	1632	CCACCTTGAGCACCAGACCGGTTGACAGGACCGGATGTTCAATTGAACTCAACCGCC	1691
DB	491	GCGACTGCGAGATCACTGTGCCCCGAGATCAAGTGGTGGGCACTACGAGTGCACCATCC	550
QY	1692	GTCTGAGCATGACTTTG	1709
DB	551	GCCTGCACCCCGAGTTG	568

Search completed: July 20, 2003, 15:27:45  
Job time : 7016 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 2003, 18:20:49 ; Search time 10.5714 Seconds  
(without alignments)  
2095.115 Million cell updates/sec

Title: US-09-807-802A-17  
Perfect score: 2906  
Sequence: 1 MASGGGAPMADNNEGADGVG.....NNGLYTEPRPIGTRYLTRPL 534

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2177.5	74.9	504	1 COA3_AAV2	P03135 adeno-ascoc
2	517	17.8	673	1 COAT_PAVBO	P07297 bovine parv
3	487.5	16.8	781	1 COAT_PAVHB	P07299 human parvo
4	235.5	8.1	749	1 COA2_PAVPN	P22964 porcine par
5	232.5	8.0	729	1 COA1_PAVPN	P33484 porcine par
6	232.5	8.0	729	1 COA1_PAVPN	P18546 porcine par
7	230.5	7.9	729	1 COA1_PAVPN	P52501 porcine par
8	229	7.9	587	1 COAT_PAVJ3	P36310 parvovirus
9	226	7.8	722	1 COAT_PAVHH	P03136 hamster par
10	197	6.8	716	1 COAT_PAVHH	P03137 murine min
11	196.5	6.8	718	1 COAT_PAVHH	P07302 murine min
12	194	6.7	722	1 COAT_PAVHH	P27437 mink enteri
13	194	6.7	727	1 COAT_PAVHH	P04864 feline panl
14	190	6.5	584	1 COAT_PAVC2	P30129 canine parv
15	189	6.5	584	1 COAT_PAVC2	P24840 feline panl
16	188	6.5	727	1 COAT_PAVCB	P11213 canine parv
17	188	6.5	737	1 COAT_PAVCD	P17455 canine parv
18	184	6.3	748	1 COAT_PAVCN	P12930 canine parv
19	177.5	6.1	647	1 COAT_PAVG	P24029 aleutian m
20	176	6.1	722	1 COAT_PAVC7	P04863 canine parv
21	122.5	4.2	648	1 WHN_HUMAN	O15353 homo sapien
22	121	4.2	880	1 SYV_BACTS	P11931 bacillus st
23	109	3.8	1186	1 CEAA_BACTS	P08704 klebsiella
24	108	3.7	655	1 CDGT_KLEPN	O45710 bacillus th
25	107.5	3.7	648	1 WHN_MOUSE	O61575 mus musculu
26	106.5	3.7	1070	1 Y355_HUMAN	O15063 homo sapien
27	106	3.6	1271	1 Y338_MYCGE	P47580 mycoplasma
28	105	3.6	1396	1 ITA2_DROME	P12080 drosophila
29	104.5	3.6	559	1 HNF6_PIG	O03365 sus scrofa
30	104.5	3.6	717	1 CLAI_ARATH	O38854 arabidopsis
31	104	3.6	395	1 MAFI_YEAST	P41910 saccharomyc
32	103.5	3.6	1196	1 AMVB_PAEPO	P21543 paenibacill
33	102	3.5	1122	1 ADPI_MYCGA	O49379 mycoplasma

34	102	3.5	2493	1 CYAA_USTMA	P49606 ustilago ma
35	102	3.5	3712	1 LMA_DROME	O00174 drosophila
36	100.5	3.5	531	1 HEXB_PIG	O29548 sus scrofa
37	100	3.4	739	1 CATA_MYCSM	O59557 mycobacteri
38	99	3.4	879	1 GUNI_CLOTM	O02934 clostridium
39	98.5	3.4	557	1 HNF6_RAT	P23899 rattus norv
40	97.5	3.4	560	1 GAT2_YEAST	P40209 saccharomyc
41	97.5	3.4	676	1 VGP_EBOIC	O66810 ebola virus
42	97.5	3.4	685	1 RNF6_HUMAN	O9y252 homo sapien
43	97.5	3.4	1113	1 N116_YEAST	O02630 saccharomyc
44	97.5	3.4	2233	1 COAC_YEAST	O00955 saccharomyc
45	97	3.3	800	1 GUN_BACSI	P06564 bacillus sp

ALIGNMENTS

RESULT 1	ID	COA3_AAV2	STANDARD;	PRT;	504 AA.
AC	P03135;				
DT	21-JUL-1986 (Rel. 01, Created)				
DT	21-JUL-1986 (Rel. 01, Last sequence update)				
DT	01-NOV-1991 (Rel. 20, Last annotation update)				
DE	Probable coat protein 3.				
OC	Adeno-associated virus 2 (AAV2).				
OC	Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.				
OX	NCBI_TaxID=10804;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=83164299; PubMed=6300419;				
RA	Srivastava A., Lusby E.W., Berns K.L.;				
RT	"Nucleotide sequence and organization of the adeno-associated virus 2				
RT	genome.";				
RL	J. Virol. 45:555-564(1983).				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@ib-sib.ch">license@ib-sib.ch</a> ).				
CC	-----				
DR	EMBL; J01901; AAA42376.1; -.				
DR	PIR; A03698; VCPV3A.				
DR	InterPro; IPR001403; Parvo_coat.				
DR	Pfam; PF00740; Parvo_coat; 1.				
KW	Coat protein.				
SQ	SEQUENCE 504 AA; 56366 MW; 75899B017052B6B2 CRC64;				
Query Match	74.9%; Score 2177.5; DB 1; Length 504;				
Best Local Similarity	82.5%; Pred. No. 5.7e-149;				
Matches 392; Conservative 33; Mismatches 47; Indels 3; Gaps 2;					
OY	1 MASGGGAPMADNNEGADGVGASGNWCHDSTWLGDRVITTSRTWALPTYNNHLYKQISS 60				
DB	1 MATSGGAPMADNNEGADGVGSSGNWCHDSTWMDRVITTSRTWALPTYNNHLYKQISS 60				
OY	61 ASTGASNDNHFGYSTPMGYEDNRFCHFSPPRDWQRLNNNWGFPRKRLNFKLFNIQVK 120				
DB	61 QS-GASNDNHFGYSTPMGYEDNRFCHFSPPRDWQRLNNNWGFPRKRLNFKLFNIQVK 119				
OY	121 EVTTNDGVTTIANNLTSTVOFSSEYQLPYVLGSAHOGCLPPFPADVFMIPQYGLTLN 180				
DB	120 EVTQNDGTTTIANNLSTVOFTDSEYQLPYVLGSAHOGCLPPFPADVFMIPQYGLTLN 179				
OY	181 NGSQAVGRSSFYCLEYFPQWLRGNNFTFSYTFEEVPHSSYAHQSQSLDRLMPLIDQY 240				
DB	180 NGSQAVGRSSFYCLEYFPQWLRGNNFTFSYTFEDVPHSSYAHQSQSLDRLMPLIDQY 239				
OY	241 LYLINRTQNSGSAQNKLDFSRGSPAGMSVQPKNWLPGPCYRQQRVSKTKTDNNNSNFT 300				



Db 240 LYYLSRTNTPSGTTQSRLOFSQAGASDIRDQSRNWLPGPCYRQQRVSKTSADNNNSEYS 299  
QY 301 WTGASKYNLNGRESINPGTAMASHKDEDEKFPMSGVMIFGKESAGASNTALDNVMTD 360  
Db 300 WTGATKYHLNGRDSLVP--AMASHKDEDEKFPQSGVLIFFGKQSEKTNVNI EKVMITD 357  
QY 361 EEEIKATNPVATERRGTVAVNFQSSSTDPAQDVHAMGALPGMVMQDRDYYLQGPWAKI 420  
Db 358 EEEIGTNPVATEQYGSVSTNLQGRNQAATADYNTQGVLPGMVMQDRDYYLQGPWAKI 417  
QY 421 PHTDGHFHPSPLMGCGFLKNPPOILIKNTPVANPAPAEFSATKFAFITQYSTG 475  
Db 418 PHTDGHFHPSPLMGCGFLKNPPOILIKNTPVANPSTTFSAAKFAFITQYSTG 472

RESULT 2

COAT\_PAVBO STANDARD; PRT; 673 AA.  
ID COAT\_PAVBO STANDARD; PRT; 673 AA.  
AC P07297; Q84374;  
DT 01-APR-1988 (Rel. 07, Created)  
DT 01-APR-1988 (Rel. 07, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Coat protein VP2 [Contains: Coat protein VP3].  
OS Bovine parvovirus (BPV).  
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.  
OX NCBI\_TaxID=10784;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=87061184; PubMed=3783814;  
RA Chen K.C., Shull B.C., Moses E.A., Lederman M., Stout E.R.,  
RA Bates R.C.;  
RT "Complete nucleotide sequence and genome organization of bovine parvovirus."  
RL J. Virol. 60:1085-1097(1986).  
CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
DR EMBL; M14363; AAB59848.1; -.  
DR EMBL; M14363; AAB59849.1; -.  
DR PIR; A26104; VCPVB5.  
DR HSSP; P30129; 4DPV.  
DR InterPro; IPR001403; Parvo\_coat.  
DR Pfam; PF00740; Parvo\_coat; 1.  
KW Coat protein; Glycoprotein.  
FT CHAIN 1 673 COAT PROTEIN VP2.  
FT CARBOHYD 138 673 COAT PROTEIN VP3.  
FT CARBOHYD 343 343 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 344 344 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 446 446 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 639 639 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT DOMAIN 163 187 GLY-RICH.  
SQ SEQUENCE 673 AA; 75103 MW; 5F244642B2214831 CRC64;

Query Match 17.8%; Score 517; DB 1; Length 673;  
Best Local Similarity 25.1%; Pred. No. 1.2e-29;  
Matches 143; Conservative 85; Mismatches 208; Indels 134; Gaps 17;

QY 5 GGAPMADNNEGADGVGNASGNMHCSTWLGDVITTTSTRTWALPTYNHLYKQISSASTG 64  
Db 163 GSGSVGGGGRGSGYSTGTGTFISENIVTKNTRQFICDIKNGHLYKS-EVLNTG 221  
QY 65 ASNDNHFGYSTPGYFDPNRFCHFSPRDQRLINNMGFRPKRLNFKLFNIQVKEVTT 124  
Db 222 DTAHRQY-AITTPSYFNFNQYSSHSFSPNDWQHLVNDYERFRPKAMIVRYNLQIKQIMT 280  
QY 125 NDGVTTIANN-LTSTVQVFSDESEYQLPYVLGSAHQGLPPFPADVFMIPOYGYLTL---- 179

Db 281 DGAMGYTNNDLTAGMHIFCDGDHRYVQHPWDQCMPELNSIWELLPOYATIPAPISV 340  
QY 180 --NNGSQAVGRS-----SFYCLEYRPSQMLRTGNNFTFSYTFEEVFPHSSYAHSQSLDR 232  
Db 341 VDNNTNTVBEHLKGVPLMYLENSDHEVLNG-----RI 375  
QY 233 MNPLIDQYLYLNRTON-----QSGSAQNKDLLFSRGP-----AGMSVQPKMWL 277  
Db 376 YRIYIQWLRLMDRKQHHIQHASDVQSTGQKQKLLIQRTQPNKQRFQNAALRTSNMW 435  
QY 278 PGPCYRQQRVSKTKTDNNNSNFTWTGASKYNLNGRESINPGTAMASHKDEDEKFPMSG 337  
Db 436 SGP-----GIARCTHNATLQTSAGALVTMTNGAD-----VSG 469  
QY 338 VM-----IFGKE-----SAGASNTALDNVMTDEEIKATNPVAT 372  
Db 470 VRAVRVGYSTDPYIGQQPESDLLRLRYSASABEQQNPLEN-----AAR 515  
QY 373 ERFGTAVNFQSSSTDPAQDVHAMGALPGMVMQDRDYYLQGPWAKIPHTDGHFHPSP 432  
Db 516 HTFTREARTKLITGSGADGDYKEWMWLPNQWDSAPISRYNPVWKVPRVNRKTLDTQ 575  
QY 433 MGFGLKNPPOILIK--NTPVANPAPAEFSATKFAFITQYSTGOVSVEIEMELQKENS 490  
Db 576 DGSIPMSHPDGTIFIKLARIIPVPGND-----SFLNIYVTGOVSCEVWEVEKRG 626  
QY 491 KRWNPVOYTSNYAKSANVD-FTVDNNGLY 519  
Db 627 KNWRPEYMHs--ATNMSVDAYTINNAGVY 653

RESULT 3

COAT\_PAVHB STANDARD; PRT; 781 AA.  
ID COAT\_PAVHB STANDARD; PRT; 781 AA.  
AC P07299;  
DT 01-APR-1988 (Rel. 07, Created)  
DT 01-APR-1988 (Rel. 07, Last sequence update)  
DT 01-APR-1990 (Rel. 14, Last annotation update)  
DE Probable coat protein VP1.  
OS Human parvovirus B19.  
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Erythrovirus.  
OX NCBI\_TaxID=10798;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Isolate AU;  
RX MEDLINE=86200451; PubMed=3701931;  
RA Shade R.O., Blundell M.C., Cotmore S.F., Tattersall P., Astell C.R.;  
RT "Nucleotide sequence and genome organization of human parvovirus B19 isolated from the serum of a child during aplastic crisis."  
RL J. Virol. 58:921-936(1986).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
DR EMBL; M13178; AAA66867.1; -.  
DR PIR; A24299; VCPV19.  
DR InterPro; IPR001403; Parvo\_coat.  
DR Pfam; PF00740; Parvo\_coat; 1.  
KW Coat protein; Glycoprotein.  
FT CARBOHYD 46 46 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 184 184 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 220 220 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 293 293 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 781 AA; 86015 MW; 8C6254DBD0576B07 CRC64;

Query Match 16.8%; Score 487.5; DB 1; Length 781;  
Best Local Similarity 27.4%; Pred. No. 2e-27;

```

Matches 143; Conservative 84; Mismatches 239; Indels 55; Gaps 16;

QY 10 ADNNEGADVGNAS--GNMHDSTWLGDRVITTTSTRTVALPTVNNHLYKQISSASTGASN 67
   | : | | | | | | | | | : | : | : | : | : | : | : | : | : | : |
Db 234 AEASTGAGGGGSSSVKSMWSEGAFTFSANSVTCFTFSRQFLPYDPEHHYKVFSPAASCHN 293
QY 68 D-----NHYFGYSTPWGYFDENRFHCHFSPRDQRLINNNGFRRPKRLNFKLENI 117
   | : | | | | | | | | | | | : | : | : | : | : | : | : | : | : |
Db 294 ASGEAKVCTISPIMGYSTFWRYLDENALNFFSPLFEQHLENYGSIAPDALVTITSEI 353
QY 118 QVKEVT--TNDGVTTIANNLSTVQVFSDESEYQLPYVLGSAHQCLPPFPADVFMIPQY 175
   | : | | | | | | | | : | : | : | : | : | : | : | : | : | : |
Db 354 AVKDVTDKTGGV-QVTDSTTGRLLCMLVDHEKYKPYVLGGQODTLAPELPIWVFFPQYA 412
QY 176 YLTINN-GSQAVG-----RSSFCLEYFPSSQMLRTGNNTFSYTFEEVPFHSYAH 225
   | | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 413 YLTVGDVNTQGISGDSKKLASEESAFAFYVLEHSSFQLLGTTASMSYKFPVPVPPENLEG 472
QY 226 SQSLDRMLNPLIDQYLYLNKTQNOGSAQNKDLLFSRGSFAGMSVQPKMLPGPCYRQ 285
   | | | | | | | | | | | | | | | : | : | : | : | : | : | : |
Db 473 SQHFYEMYNPL--YGSRLGVPTLGGDPKFRSL----THEDHAIQPNFMPLVNSV 524
QY 286 RVSKTKTDNNNSNFTWTGASKYNNLNGRESIINPG-TAMASHKDEDEKFFPMGSMVIFGKE 344
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 525 STKEGSSNTGAKKALTGLSTGTSQNTTRISLRPGPVSPYHHWDTDKYVTGINAISHQT 584
QY 345 SAGASNTALDNV-----MITDEEEKATNPVATERFGTAVANFQSSSTDPATGVHAM 397
   | : | | | | | | | : | : | : | : | : | : | : | : | : | : |
Db 585 TYG--NAEDKEYQGGVGRFPNEKEQLKQLGLNMHTY-----FPNKGTOQYTDQIE-R 634
QY 398 GALPGMWQDRDVLQGPIMAKIPHTDGHFHS-PLMGFGFLKNPPQILIKNTPVPANP 456
   | : | | | | | | | : | : | : | : | : | : | : | : | : | : |
Db 635 PLMWGSVMNRRALHYESQLSMKIPNLDDSFKTQFALGGMGLHQPPQIFLK--ILPQSG 692
QY 457 PAEFSATKFAFITQYSTGOVSVEIEWEL-QKENSKRNNDE 496
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 693 PIGGIKSMGITTLVQYAVGIMVTMTPKLGPRAKATGRMNPQ 733

RESULT 4
COA2_PAVPN STANDARD; PRT; 749 AA.
ID _COA2_PAVPN
AC P22964;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Coat protein VP1 [Contains: Coat protein VP2] (Version 2).
OS Porcine parvovirus (strain NADU-2) (PPV).
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
OX NCBI_TaxID=10797;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=90010964; PubMed=2794971;
RX Ranz A.I., Mancius J.J., Diaz-Aroca E., Casal J.I.;
RT "Porcine parvovirus: DNA sequence and genome organization.";
RL J. Gen. Virol. 70:2541-2553(1989).
CG -1- SUBUNIT: MATURE VIRION CONTAINS THREE CAPSID PROTEINS DESIGNATED
   VP1, VP2, AND VP3 AND A NONCAPSID PROTEIN NS-1.
CC -1- MISCELLANEOUS: VP3 MIGHT BE A POSTTRANSLATIONAL CLEAVAGE PRODUCT
   OF VP2 IN SEVERAL AUTONOMOUS PARVOVIRUSES.
CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, D00623; BAA00502.1; -.
DR PIR; B33302; VCPVPP.
DR HSSP; P30129; 4DPV.

```

DR InterPro; IPR001403; Parvo\_coat.  
 DR Pfam; PF00740; Parvo\_coat; 1.  
 KW Coat protein; Glycoprotein.  
 FT CHAIN 1 749 COAT PROTEIN VP1.  
 FT CHAIN 171 749 COAT PROTEIN VP2.  
 FT DOMAIN 195 210 POLY-GLY.  
 FT CARBOHYD 192 192 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 218 218 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 302 302 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 350 350 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 453 453 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 491 491 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 593 593 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 624 624 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 671 671 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 749 AA; 82872 MW; 6CA13CB97BE93418 CRC64;  
 Query Match 8.1%; Score 235.5; DB 1; Length 749;  
 Best Local Similarity 23.2%; Pred. No. 2.3e-09;  
 Matches 139; Conservative 83; Mismatches 241; Indels 137; Gaps 32;  
 QY 3 SGGGAPMADNNEGADGVGNASGNMH--CDSTWLGD---RVITSTRTVALPTYNNHLVXQ 57  
 DB 194 SGGGGG-GGGGRGAGVGSTGSFNNQTEFOYLGELVRITAHASRLIHLNMPHEHYKR 252  
 QY 58 I-----SSASTGAS-NDNHFGYSTPWGYFDENRFHCHSPRDQRLINNNGFRPKRLN 111  
 DB 253 IHLNSESAGQWQDDAHTQMTVPMSLIDANAWGVENPADWQLISNNMTEINLVSE 312  
 QY 112 FKLFNIOQKEVT--TNDGVTIANMLTSTVOVFSDEXQLPYVLGSAHQGLPPFPADV 168  
 DB 313 QALFNVLKTTESATSPPTKIYNNDLTASLWALDTNNTLPYTPAPRSETTLGFYFWLP 372  
 QY 169 FMIPQGYL-----TLNNGSQAVGRS-----SFYCLE-YFPSQMLRTGNF 208  
 DB 373 TKPTQYRYLLSCIRLNLPPTYTGQSQITDSIQTGLHSDIMEFYTIENAVPIHLRTGDEF 432  
 QY 209 -TFSTFEFVPHSSVAHSQSLDR-----LNNPLL--DQYLYLNRTQNGSQAQNK 257  
 DB 433 STGIYHFDTKPL--KLTHSWQTNRSGLPKLLEPTTEGDQHPGTLPAANTRKGYHOTI 490  
 QY 258 DLLFSRGPAGMSVQDKNMLPGPCYRQQRVSKTKTDNNNSNFTWTGASKYNLNGRESTIN 317  
 DB 491 NNSYTEAT---AIRP-----AQVGINTPYMNFYSNGGP-----LT 524  
 QY 318 P--GTAMASHKDDDEKFFPMGVMIFGKESAGASNTALD--NVMITDEEEK--ATNP- 369  
 DB 525 PIVPTADTQYNDDEPN-----GAIRFTMDYQHGHLTTSSQELERYTFNPQ 569  
 QY 370 ----VATERFGTVA--VNFQSSS-----TDPATG--DVHAMGALP----GMWQDRDV 410  
 DB 570 SKCGRAPKQGFNQCAPLNLNTNNGTLPSDPIGKSNMHFMNTLNTYGPLTALNNTAPV 629  
 QY 411 YLQGPWAKIPIHTD--GHFHPSPLMGFGLK-NPPQQLIKNTVPANPPAFBSA-TKFA 466  
 DB 630 FPNQGIWMDKELDTDLKPRLLH---VTAPFVCKNNPPGOLFVKIAP---NLTDDEFNADSPQ 683  
 QY 467 SFITQYSTQGVSVIEIMELQKENSKRWNPEVOYTSNYAKSANVDFTVDDNGLYTEPRPIG 526  
 DB 684 PRITTSNPFMWKGTILFTAKMRSSNMMPIQOHTT-----TAENIGKYI-PTNIG 732  
 RESULT 5  
 COAL\_PAVP9  
 ID COAL\_PAVP9 STANDARD; PRT; 729 AA.  
 AC P33484;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Coat protein VP1 [Contains: Coat protein VP2].  
 OS Porcine parvovirus (strain 90HS) (PPV).  
 OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.  
 NCBI TaxID=33725;

```

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89319168; Pubmed=2750278;
RA Sakurai M., Nishimori T., Ushimi C., Nakajima H.;
RT "Nucleotide sequence of capsid protein gene of porcine parvovirus.";
RL Virus Res. 13:79-86(1989).
CC -1- SUBUNIT: MATURE VIRION CONTAINS THREE CAPSID PROTEINS DESIGNATED
CC VP1, VP2, AND VP3 AND A NONCAPSID PROTEIN NS-1.
CC -1- MISCELLANEOUS: VP3 MIGHT BE A POSTTRANSLATIONAL CLEAVAGE PRODUCT
CC OF VP2 IN SEVERAL AUTONOMOUS PARVOVIRUSES.
CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.
DR PIR; A60006; A60006.
DR HSSP; P30129; 4DPV.
DR InterPro; IPR001403; Parvo coat.
DR Pfam; PF00740; Parvo_coat; 1.
KW Coat protein; Glycoprotein.
FT CHAIN 1 729 COAT PROTEIN VP1.
FT CARBOHYD 151 729 COAT PROTEIN VP2.
FT CARBOHYD 172 729 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 198 198 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 282 282 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 433 433 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 471 471 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 604 604 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 651 651 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 729 AA; 80938 MW; 92538BFE9A0C78E6 CRC64;

Query Match 8.0%; Score 232.5; DB 1; Length 729;
Best Local Similarity 23.0%; Pred. No. 3.6e-09;
Matches 137; Conservative 86; Mismatches 245; Indels 127; Gaps 31;

QY 3 SGGAPMADNNEGADVGNASGNMH--CDSTWLGD--RVITSTRTWALPTYNNHLYKQ 57
DB 174 SGGGGG-GGGGRGAGGVSTGSFNNQTEFOYLGEGLVRLTAHASRLIHLNMPHEHYKR 232
QY 58 IS--SASTGAS---NDNHYFGYSTPWGYFDENRFCHFSPRDWQRLINNNWGFRPKRLN 111
DB 233 IHVLNSESAGVQWQDDAHTQMTVPWSLIDANAMGWENPADWQLISNNMTEINLVSEFE 292
QY 112 FKLFNIOQKEVT--TNDGVTTIANNTSTVOYFSDSEYQLPYVLGSAHQCLPPPADV 168
DB 293 QEIFNVVLKTTTESATSPPTKIYNNDLTASLVALDTNNTLPYTPADPSETLGFYPMWP 352
QY 169 FMIPQGYL-----TLNNGSQAVGRS-----SFYCLE-YFPSQMLRTGNNE 208
DB 353 TKPTQYRYLSCTRNLNPTTYTGQSQQITDSIQGLHSDIMEFYTIENAVFIHLRTGDEF 412
QY 209 -TFSYTFEEVPEHSSYASQSLSR-----IMNPLI--DQYLYYINRTONQSGSAQNK 257
DB 413 STGIYHFDTKPL--KLTHSWQTNRLGLPKLLTEPTTEGDQHPGLPAANTRKGXHTM 470
QY 258 DLLFSRGSFAGMSVQPKNMLPGPCYRQQRVSKTKTDNNNSNFTWTGASKYNLNGRESIIN 317
DB 471 NNSYTEAT-----AIRP-----AQVGYNTPYMNFEYSNGGPF-----LT 504
QY 318 P--GTAMASHKDEDEKFFPMSCVMIFGKESAGASNTALDNVMTDEEIKATNP----- 369
DB 505 PIVPTADTYQNDDE---PNGAIRFTMGYHQGLTTS-----SQELERYTFNPQSKCGR 554
QY 370 VATERFETVA-VNFQSSS-----TDPATG--DVHAMGALP-----GMWQDRDVLQGP 415
DB 555 APKQOFNQAPLNLNENTNNGTLPLSPDPIGKPMHFMNTLNTYGPLTALNNTAPVFPNGQ 614
QY 416 IWAKIPIHTD--GHFHPSPMLMGFGLK--NPPQILIKNTPVPAHPAEFSA-TKFAFITQ 471
DB 615 IWDKELDTDLKPRLLH--VTAPFVCKKNPPQQLFVKIAP--NLTDENADSPQOPRIIT 668
QY 472 YSTGQVSVEIEWELQKENSQRWNPEVOYTSNYAKSANDFTVDNNGLYTEPRPIG 526
DB 669 YSNFWWKGTLFTAKRRSSNMWNPICQHTT-----TAENIGNYI-PTNIG 712

```

```

RESULT 6
ID_COAL_PAVPN STANDARD; PRT; 729 AA.
AC P18546; Q89816;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Coat protein VP1 [contains: Coat protein VP2] (Version 1).
OS Porcine parvovirus (strain NADL-2) (PPV).
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
OX NCBI_Taxid=10797;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91021005; Pubmed=2219713;
RA Vasudevacharya J., Basak S., Srinivas R.V., Compans R.W.;
RT "The complete nucleotide sequence of an infectious clone of porcine
RL parvovirus, strain NADL-2.";
RL Virology 178:611-616(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90085785; Pubmed=2596019;
RA Vasudevacharya J., Basak S., Srinivas R.V., Compans R.W.;
RT "Nucleotide sequence analysis of the capsid genes and the right-hand
RL terminal palindromic sequence of porcine parvovirus, strain NADL-2.";
RL Virology 173:368-377(1989).
CC -1- SUBUNIT: MATURE VIRION CONTAINS THREE CAPSID PROTEINS DESIGNATED
CC VP1, VP2, AND VP3 AND A NONCAPSID PROTEIN NS-1.
CC -1- MISCELLANEOUS: VP3 MIGHT BE A POSTTRANSLATIONAL CLEAVAGE PRODUCT
CC OF VP2 IN SEVERAL AUTONOMOUS PARVOVIRUSES.
CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M38367; AAA46919.1; -.
DR EMBL; M38367; AAA46921.1; -.
DR EMBL; M32787; AAA46917.1; -.
DR EMBL; M32787; AAA46918.1; -.
DR PIR; B33743; VCPVNA.
DR HSSP; P30129; 4DPV.
DR InterPro; IPR001403; Parvo coat.
DR Pfam; PF00740; Parvo_coat; 1.
KW Coat protein; Glycoprotein.
FT CHAIN 1 729 COAT PROTEIN VP1.
FT CARBOHYD 151 729 COAT PROTEIN VP2.
FT CARBOHYD 172 729 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 198 198 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 282 282 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 433 433 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 471 471 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 604 604 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 651 651 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT DOMAIN 175 190 GLY-RICH.
SQ SEQUENCE 729 AA; 80946 MW; EF816E246C80DB42 CRC64;

Query Match 8.0%; Score 232.5; DB 1; Length 729;
Best Local Similarity 23.0%; Pred. No. 3.6e-09;
Matches 138; Conservative 85; Mismatches 240; Indels 137; Gaps 32;

QY 3 SGGAPMADNNEGADVGNASG--NWHCDSTWLGD--RVITSTRTWALPTYNNHLYKQ 57
DB 174 SGGGGG-GGGGRGAGGVSTGSFNNQTEFOYLGEGLVRLTAHASRLIHLNMPHEHYKR 232
QY 58 IS--SASTGAS---NDNHYFGYSTPWGYFDENRFCHFSPRDWQRLINNNWGFRPKRLN 111

```



```
Db 233 IHVLNSESQVAGQMVQDDAHTQMTWTPWSLIDANAMGWENPADWQLISNNMTEINLVSE 292
Qy 112 EKLFNIOQKEVT---TNDGVTIANNLSTVQVFSDEYQLPYVLGSAHQGLPPFPADV 168
Db 293 QEIFNVVLKLTITESATSPPTKIYNNDLTASIMVALDTNNTLPTTPAARSETLGFYFWLP 352
Qy 169 FMIPQYGYL-----TLNNGSQAVGRS-----SFYCLE-YFPSQMLRTGNMF 208
Db 353 TKPTQYRYLLSCIRNLNPTTYGQSQQITDSIQTLGHSIDIMFYTTENAVPIHLRTGDEF 412
Qy 209 -TFSYTFEEVPEFHSSYAHQSQSLDR-----LMNPLI--DQYLYLNRTQNGSQAQNK 257
Db 413 STGIYHFDTKPL--KLTHSWQTRNSLGLPKLLEPTTEGDQHPGTLPAANTRKGYHQT 470
Qy 258 DLLFSRGSFAGMSVQPKNWLPGPCYRQQRVSKTKTDNNNSNFTWTGASKYNLNGRESIIN 317
Db 471 NNSYTEAT-----AIRP-----AQVGYNTPYMNFYSNGGPF-----LT 504
Qy 318 P--GTAMASHKDDDEKFPFMSGVMIFGKESAGASNTALD--NVMITDEEEK--ATNP- 369
Db 505 PIVPTADTQYNDDEPN-----GAIRFTMDYQHGHLTSSQELERYTFNPQ 549
Qy 370 -----VATERFGTVA-VNFQSSS-----TDPATG--DVHAMGALP-----GMWQDRDV 410
Db 550 SKCGRAPKQGFNQQAPLNLENTNNGTLPSDPIGKSNMHFMNTLNTYGPLTALNTAPV 609
Qy 411 YLQGPIMAKIPHTD--GHFHPSPLMGFGLK-NPPQILIKNTPEVPANPPAEFSA-TKFA 466
Db 610 FPNQGMKELDTLKPRLH---VTAPFVCKNNPPGQLFVKIAP---NLTDDEFNADSPQ 663
Qy 467 SFITQYSTGVSVIEIEMELQKENSKRNNPEVQYTSNYAKSANDVFTVDNNGLYTEPRPIG 526
Db 664 PRITTSNFWWKGTLTFTAKMRSSNMNNDPIQHTT-----TAENIGNYI-PTNIG 712
```

```
RESULT 7
COAT_PAVPK
ID COAT_PAVPK STANDARD; PRT; 729 AA.
AC P52501;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Coat protein VP1 [Contains: Coat protein VP2].
OS Porcine parvovirus (strain Kresse) (PPV).
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
OX NCBI_TaxID=73487;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96183900; PubMed=8642680;
RA Bergeron J., Hebert B., Tijssen P.;
RT "Genome organization of the Kresse strain of porcine parvovirus:
RT identification of the allotropic determinant and comparison with
RT those of NADL-2 and field isolates.";
RL J. Virol. 70:2508-2515(1996).
CC -|- SUBUNIT: MATURE VIRION CONTAINS THREE CAPSID PROTEINS DESIGNATED
CC VP1, VP2, AND VP3 AND A NONCAPSID PROTEIN NS-1.
CC -|- MISCELLANEOUS: VP3 MIGHT BE A POSTTRANSLATIONAL CLEAVAGE PRODUCT
CC OF VP2 IN SEVERAL AUTONOMOUS PARVOVIRUSES.
CC -|- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
CC -----
DR EMBL; U44978; AAC40230.1; -.
DR EMBL; U44978; AAC40231.1; -.
DR HSSP; P30129; ADPV.
DR InterPro; IPR001403; Parvo_coat.
DR Pfam; PF00740; Parvo_coat; 1.
```

```
KW Coat protein; Glycoprotein.
FT CHAIN 1 729 COAT PROTEIN VP1.
FT CHAIN 151 729 COAT PROTEIN VP2.
FT CARBOHYD 172 172 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 198 198 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 282 282 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 433 433 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 471 471 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 604 604 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 651 651 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 729 AA; 80835 MW; B6345BFA0568A1F6 CRC64;
```

Query Match 7.9%; Score 230.5; DB 1; Length 729;  
Best Local Similarity 23.0%; Pred. No. 5e-09;  
Matches 137; Conservative 86; Mismatches 245; Indels 127; Gaps 31;

```
Qy 3 SGGGAPMADNNEGADGVGNAGNWH--CDSTWLGD--RYITSTRTWALPTYNNHLYKQ 57
Db 174 SGGGGG-CGGGRGAGGVGVSFGSNNQTEFYQLGGLVRIITAHASRLIHLNMPHEHYKR 232
Qy 58 IS--SASTGAS--NDNHYFGYSTPWGYFDNRFHCHSPRDWQRLINNNWGFPRKLN 111
Db 233 IHVLNSESQVAGQMVQDDAHTQMTWTPWSLIDANAMGWENPADWQLISNNMTEINLVSE 292
Qy 112 EKLFNIOQKEVT---TNDGVTIANNLSTVQVFSDEYQLPYVLGSAHQGLPPFPADV 168
Db 293 QEIFNVVLKLTITESATSPPTKIYNNDLTASIMVALDTNNTLPTTPAARSETLGFYFWLP 352
Qy 169 FMIPQYGYL-----TLNNGSQAVGRS-----SFYCLE-YFPSQMLRTGNMF 208
Db 353 TKPTQYRYLLSCIRNLNPTTYGQSQQITDSIQTLGHSIDIMFYTTENAVPIHLRTGDEF 412
Qy 209 -TFSYTFEEVPEFHSSYAHQSQSLDR-----LMNPLI--DQYLYLNRTQNGSQAQNK 257
Db 413 STGIYHFDTKPL--KLTHSWQTRNSLGLPKLLEPTTEGDQHPGTLPAANTRKGYHQT 470
Qy 258 DLLFSRGSFAGMSVQPKNWLPGPCYRQQRVSKTKTDNNNSNFTWTGASKYNLNGRESIIN 317
Db 471 NNSYTEAT-----AIRP-----AQVGYNTPYMNFYSNGGPF-----LT 504
Qy 318 P--GTAMASHKDDDEKFPFMSGVMIFGKESAGASNTALDNVMITDEBEIKATNP----- 369
Db 505 PIVPTADTQYNDDE---PNGAIRFTMGYQHGQLTTS-----SQELERYTFNPQSKGR 554
Qy 370 VATERFGTVA-VNFQSSS-----TDPATG--DVHAMGALP-----GMWQDRDVYIQGP 415
Db 555 APKQGFNQQAPLNLENTNNGTLPSDPIGKSNMHFMNTLNTYGPLTALNTAPVFNQ 614
Qy 416 IWAKIPHTD--GHFHPSPLMGFGLK-NPPQILIKNTPEVPANPPAEFSA-TKFAPIQ 471
Db 615 IWDKELDTLKPRLH---VTAPFVCKNNPPGQLFVKIAP---NLTDDEFNADSPQPRIT 668
Qy 472 YSTGVSVIEIEMELQKENSKRNNPEVQYTSNYAKSANDVFTVDNNGLYTEPRPIG 526
Db 669 YSNFWMWKGTLTFTAKMRSSNMNNDPIQHTT-----TAENIGNYI-PTNIG 712
```

```
RESULT 8
COAT_PAVL3
ID COAT_PAVL3 STANDARD; PRT; 587 AA.
AC P36310;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Coat protein VP1 [Contains: Coat protein VP2].
OS Parvovirus Iu111.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
OX NCBI_TaxID=35339;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93297126; PubMed=8517025;
```

RA Diffot N., Chen K.C., Bates R.C., Lederma M.;  
RT "The complete nucleotide sequence of parvovirus Lu11 and  
RT localization of a unique sequence possibly responsible for its  
RT encapsidation pattern."  
RL Virology 192:339-345(1993).  
CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M81888; -; NOT\_ANNOTATED\_CDS.  
DR PIR; B44276; B44276.  
DR HSSP; P07302; 1MVM.  
DR InterPro; IPR001403; Parvo\_coat.  
DR Pfam; PF00740; Parvo\_coat; 1.  
KW Coat protein; Glycoprotein.  
FT CARBOHYD 49 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 90 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 220 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 304 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 371 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 503 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 511 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 514 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 539 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 571 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 587 AA; 65429 MW; 523E6B9CBF2EBB74 CRC64;

Query Match 7.9%; Score 229; DB 1; Length 587;  
Best Local Similarity 22.7%; Pred. No. 4.8e-09;  
Matches 133; Conservative 74; Mismatches 228; Indels 150; Gaps 28;

QY 3 SGGGAPMADNNEGADGVGNASGNHCHDS--TWLGD---RVITSTRTWAL--PTYNN--- 52  
DB 30 SGGG-----GSGGGGVGVSTGSYDNQTHYKFLGDGWEITAYSTRMVLNMEKSENYCR 83  
QY 53 ---HLVKQISSASTGASNDNHYFGYSTPWGYFDENRFCHFP3PRDWQRLINNNGFRPKR 109  
DB 84 VRVHNTNDGTASHMAMDAHEQIW-TPWSLVDANAMGWVFPQSDWQYISNMHINLHS 142  
QY 110 LNFKLFNIOQKEVT-TNDGVTTIA--NNLTSTVOYFSDSEYQLPYVLGSAHQGL--- 161  
DB 143 LDQELFNVIKTVTEQNTGAEAIKVYNNDLTAAMVVALDSNNILPYTPAIDNQETLGFYP 202  
QY 162 --PPFPA-----DVFMIPOGYLTLNNGSOAVGRSSFYCLEYFPPS--QMLRTGN 206  
DB 203 WKPTIPSPRYFYFSCDRNLSTYKDEAGTITDTWGLASGLN3QFTIENTFORINLLRTGD 262  
QY 207 NF-TESYTFEEVPHFSYAHQSGLRLMPLIDQ-----LYLYNRTQN- 249  
DB 263 EYATGTYFDTEPIRLTHWTQNRHLGQRPQITELPSSDTANATLTARGYRSGLTQIQGR 322  
QY 250 -----QSGSAQNKD-LLFSRGPAGMSVQPKNWLPG-----PCYRQQR 286  
DB 323 NDVTETATRVRAQVGFQCPDHNFETSRAGPFKVPVVPADITQGLDHDANGSLRYTYDKQH 382  
QY 287 VSKTKTDNNNSNFTWTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMGVMIFGKESA 346  
DB 383 GQSWASQNNKDRYTW-DAVNYD-SGR----- 406  
QY 347 GASNTALDNVMTDEEIKATNPVATER--FGTAVNFQS--SSTDPAIGVHAMGALP 401  
DB 407 WTNNCFIQSVFTSEPN--ANQILTRNDLAGKTDIHTNAFNSYGLTAPFH----- 457  
QY 402 GMWVQDRDVLQGPIMAK--IPHTDGHFHPSPILMGFGFLKNPPQILIKNTP--VPANP 456  
DB 458 -----PARIYPOGQIWDKEULDEHKPRLHTQAPFV---CKNNAPGQLVRLAPNULTDQYD 509

QY 457 PAEFSATKFAFITOYSTGQVSEIEMELOKENSKRMNPEVQYTS 501  
DB 510 PNSSNLRIIVTYGTFFWKGLTLKAK--MRPNA-TWNPVQFQISA 550

RESULT 9  
COAT\_PAVHH  
ID COAT\_PAVHH STANDARD; PRT; 722 AA.  
AC P03136;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Coat protein VP1 [Contains: Coat protein VP2].  
OS Hamster parvovirus H1.  
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.  
OX NCBI\_TaxID=10799;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=83112183; PubMed=6823009;  
RA Rhode S.L. III, Paradiso P.R.;  
RT "Parvovirus genome: nucleotide sequence of H-1 and mapping of its  
RT genes by hybrid-arrested translation."  
RL J. Virol. 45:173-184(1983).  
CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X01457; CAB57285.1; ALT\_SEQ.  
DR PIR; A03699; VCPV12.  
DR HSSP; P07302; 1MVM.  
DR InterPro; IPR001403; Parvo\_coat.  
DR Pfam; PF00740; Parvo\_coat; 1.  
KW Coat protein; Glycoprotein.  
FT CHAIN 1 722 COAT PROTEIN VP1.  
FT CHAIN 131 722 COAT PROTEIN VP2.  
FT CARBOHYD 178 178 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 218 218 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 639 639 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 647 647 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 675 675 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT DOMAIN 155 170 GLY-RICH.  
SQ SEQUENCE 722 AA; 79737 MW; 6BB678391AA5DC31 CRC64;

Query Match 7.8%; Score 226; DB 1; Length 722;  
Best Local Similarity 21.7%; Pred. No. 1e-08;  
Matches 128; Conservative 83; Mismatches 234; Indels 146; Gaps 26;

QY 2 ASGGAPMADNNEGADGVGNASGNHCHDSTW--LGDRTVITSTRTWAL----- 47  
DB 153 ADGGGS-SGGGGSGGGIGVSTGYDNQTHYKFLGDGWEITAHASRLHLGMPSENYC 211  
QY 48 --PTYNNHLVKQISSASTGASNDNHYFGYSTPWGYFDENRFCHFP3PRDWQRLINNNGWF 105  
DB 212 RVTVAHNNQTTGHGTVKGNMAYDTHQOIW-TPWSLVDANAMGWVFPQSDWQFIQNSMESL 270  
QY 106 RPKRLNFKLFNIOQKEVT-----NDGVTTIANNLSTVOYFSDSEYQLPYVLGSAHQGC 160  
DB 271 NLDSLQELFNVIKTVTEQNGAGQDAIKVYNNDLTACMMVVALDSNNILPYTPAQTSET 330  
QY 161 L-----PPFPAD--VFMIPQGYLTLNNGSQ-----AVG-----RSSFYCLE-YFPPS 199  
DB 331 LGFYPPWKPTAPAPRYRYFYFMPRQLSVTSSNSAEGTQITDTIGEPQALNSQFTIENTLPI 390  
QY 200 QMLRTGNMTEF-SYTFEEVPHFSYA-----HSQSLRLMN-PLIDQYLYLNRTONQSG 252  
DB 391 TLLRTGDEFTTGYIFNTDPLKLTHTWTQNRHLACLQIGITDLEPTSDATASLTANGDRFG 450

```
QY 253 SAQNKDLF-----SRGSPAGMSVOPKMWLPGPCYRQQRVSK 289
DB 451 STQTQNVNVTYALRTRPAQIGFMQPHDNFEANRGPFKVPVVP-----LDI 497
QY 290 TKTDNNNSNFTWTSKYNLNGR--ESIINPGTA-----WASHKDEDEKFFPMGSG 337
DB 498 TAGEHDHDAN-----GAIREFNGKHGEDMAKQGAAPERYTWDALDSAGRDTCARCFV----- 549
QY 338 VMIFGKESAGASNTALDNVMTDEBEIKA-TNPVATERFGTVAVNFGSSSTDPATGDVHA 396
DB 550 -----QSAPISIPRNQOILQREDALAGRTNMHTNVFNSYGPLSAFPHBDP----- 596
QY 397 MGALPGMWQDRDVLQGPIMAK--IPHTDGHFHPSPLMGFGGLKNPPQILIKNTPPVP 453
DB 597 -----IYPNGQIWDKELDLHKPRHLHTAPFV--CKNNPPGQLFVHLGP-- 638
QY 454 ANPPAEF--SATKFAFITQYSTGVSEIEMELQKENSKRWNPEVQYTSN 502
DB 639 -NLTDQFDPNSTTVSRIVT-YSTFYWKGLIKFKAKLRPNLTWNPVYQATTD 687
```

## RESULT 10

```
COAT_MUMIV STANDARD; PRT; 716 AA.
AC P03137;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Coat protein VP1 [Contains: Coat protein VP2].
OS Murine minute virus (Murine parvovirus).
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
OX NCBI_TaxID=10794;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83143341; PubMed=6298737;
RA Astell C.R., Thomson M., Merchlinsky M., Ward D.C.;
RT "The complete DNA sequence of minute virus of mice, an autonomous
parvovirus.";
RL Nucleic Acids Res. 11:999-1018(1983).
CC -| SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC
CC EMBL; V01115; CAA24310.1; ALT_SEQ.
DR PIR; A03700; VCPV2M.
DR HSSP; P07302; 1MVM.
DR InterPro; IPR001403; Parvo_coat.
DR Pfam; PF00740; Parvo_coat_1.
KW Coat protein; Glycoprotein.
FT CHAIN 1 / 716 COAT PROTEIN VP1.
FT CHAIN 131 716 COAT PROTEIN VP2.
FT CARBOHYD 179 179 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 218 218 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 500 500 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 633 633 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 669 669 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT DOMAIN 156 171 GLY-RICH.
SQ SEQUENCE 716 AA; 78707 MW; 6A72229A91161F4C6 CRC64;
```

```
Query Match 6.8%; Score 197; DB 1; Length 716;
Best Local Similarity 20.0%; Pred. No. 1.2e-06;
Matches 126; Conservative 88; Mismatches 241; Indels 174; Gaps 28;
```

```
QY 3 SGGAPMADNNEGADGVGNASGNWHCHDS--TWLGD---RVITTSFRTWALPTYNNHLYKQ 57
DB 160 SGGG-----GSGGGGVGVSTGSDYNDQTHYRFLGDGWVEITALATRLVHLNMPKSENYCR 213
```

```
QY 58 I-----SSASTGASNDNHYFGYSTPWGYDFENRFHCHESPRDWORLINNMGFRPKRL 110
DB 214 IRVHNTDTSVKGNMAKQDAHQIWTWPSLVDANAMGVWLQPSDWQYICNTMSQLNIVSL 273
QY 111 NFKLFNIQVEKVTND---GVTTIANLSTVQVFSDEYQLPYVLGSAHQGLPPEPA 166
DB 274 DQEIFNVVLKTVTEQDLGGQAKIKYNNDLTACMVAVDSNNILPYTPAANSMETLGFYFW 333
QY 167 DVFMIPQGY-----LTNNGSQAV-----GRSSFYCLEYFPS-QMLRTGN 207
DB 334 KPTIASPFRYFYFCVDRDLSTVENQEGTVEHNVMTGPKGIPOFFTIENTQOITLRTGDE 393
QY 208 F-TPSYTFEEVFPFSSYAHQSGLDRMN--PLIDQYLYLNRTQNGSSAQNKDLFPERG 264
DB 394 FATGYTFYFDTSV--KLTHWQTNRQLGQPLLSTF---PEADTDAGT-----LTAQG 441
QY 265 SPAGMSVQPKMWL-----PGPCYRQQRVSKTKTDNNNSNFT 300
DB 442 SRHGTTQMGVNWVSEAIRTRPAQVGFQCPHNDFEASRAGP-FAAPKVPADITQGVDKAN 500
QY 301 WTGASKYNLNGRESIINPG-----TAMASHKDEDEKFF-----PMGVMIF 341
DB 501 GSVRYSYGKHGENWASHGPAERYTWDETSFGSGRDTQGFQSAPLVPPPLNGI--- 557
QY 342 GKESAGASNTALDNVMTDEEIKATNPVATERFGTVAVNFGS--SSTDPATGDVHAMGA 399
DB 558 -----LTNANPIGTKN---DIHFSNVFNSYGPLTAFSH----- 587
QY 400 LPGMWQDRDVLQGPIMAK--IPHTDGHFHPSPLMGFGGLKNPPQILIKNTPPVPA 454
DB 588 -----PSPVYPQGOIWDKELDLHKPRHLHTAPFV--CKNNAPGQMLVRLGPLNLTDQ 637
QY 455 NPPAEFSATKFAFITQYSTGVSEIEMELQKENSKRWNPEVQYTSNYAKSANVDFYVD 514
DB 638 YDPNGATLSRIVTYGTFFWKGLTMRAKLRA---NTTWNPVYQVSAA-----D 682
QY 515 NNGLV---TEPRPIGT-----RYLTRPL 534
DB 683 NGSYSVTKWLPATAGMGSVPLITRPV 711
```

## RESULT 11

```
COAT_MUMIV STANDARD; PRT; 718 AA.
AC P07302; Q9WMH3; Q9WMH2;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Coat protein VP1 [Contains: Coat protein VP2].
OS Murine minute virus (strain WVM1) (Murine parvovirus).
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
OX NCBI_TaxID=10795;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86115415; PubMed=3502703;
RA Astell C.R., Gardiner E.M., Tattersall P.;
RT "DNA sequence of the lymphotropic variant of minute virus of mice,
WVM(1), and comparison with the DNA sequence of the fibrotropic
prototype strain.";
RL J. Virol. 57:656-669(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=85242059; PubMed=3855242;
RA Sahli R., McMaster G.K., Hirt B.;
RT "DNA sequence comparison between two tissue-specific variants of the
autonomous parvovirus, minute virus of mice.";
RL Nucleic Acids Res. 13:3617-3633(1985).
```

```
RP X-RAY CRYSTALLOGRAPHY (3.5 ANGSTROMS) OF 132-718.
RA llamas-Saiz A.L., Agbandje-Mckenna M., Wikoff W.R., Bratton J.,
RA Tattersall P., Rossmann M.G.;
RT "Structure determination of Minute Virus of mice.";
RL Acta Crystallogr. D 53:93-100(1997).
```



CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; X02481; CAB46507.1; -.  
DR EMBL; X02481; CAB46508.1; -.  
DR EMBL; M12032; AAA69569.1; ALT\_INIT.  
DR PIR; B23008; VCPVIM.  
DR PDB; 1MVM; 25-FEB-98.  
DR InterPro; IPR001403; Parvo\_coat.  
DR Pfam; PF00740; Parvo\_coat; 1.  
KW Coat protein; Glycoprotein; 3D-structure.  
FT CHAIN 1 718 COAT PROTEIN VP1.  
FT CHAIN 132 718 COAT PROTEIN VP2.  
FT CARBOHYD 180 718 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 219 718 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 502 718 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 635 718 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 671 718 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT DOMAIN 157 172 GLY-RICH.  
FT CONFLICT 144 144 A -> G (IN REF. 2).  
SQ SEQUENCE 718 AA; 79017 MW; B43C1762ED5F74B9 CRC64;

Query Match 6.8%; Score 196.5; DB 1; Length 718;  
Best Local Similarity 20.5%; Pred. No. 1.4e-06;  
Matches 122; Conservative 82; Mismatches 243; Indels 147; Gaps 25;

OY 3 SGGAPMADNNEGADGVNAGNWHCD--TWLGD--RVITSTRFWALPTNNHLYKQ 57  
DB 161 SGGG-----GSGGGGVSTGSYDNQTHYRFLGDGWEITALATRLVHLNPKSENYCR 214  
OY 58 I-----SSASTGASNDNHYFGYSTPWGYFDENFRHCHFSFRDWORLNNMGFRPKRL 110  
DB 215 IRVHNTTDSVKGNMAKDAHEQIWTWPSLVDANAMGWMLQPSDMQYICNTMSQLNLVSL 274  
OY 111 NFKLFNIQKEVTND---GVTTIANNLSTVQVPSDSEYQLPYVLSAHQGLPPPPA 166  
DB 275 DQEIFNVLTQTEQDSGOAIKIYNNDLTACMVAVAVDSNNILPYTPAANSMTLGFYFW 334  
OY 167 DVFMIPQYGY-----LTINN-----GSOAVGRSSPYCLEYFPG--QMLRTGN 206  
DB 335 KPTIASPYRYFCVDRDISVTYENQEGTIEHNVMGTPKMNISQFTIENTQOITLLRTGD 394  
OY 207 NF-TFSYTFEEVPPHSSYASQSLDRLMN--PLIDQYLYLARTONQSGAONKDLLFSR 263  
DB 395 EFATGTYVFDTPNV--KLHTWQTNROLGPPILSTF---PEADTDAGT-----LTAQ 442  
OY 264 GSPAGMSVQPKNWL-----PGPCYRQQRVSKTKTDNNNSNF 299  
DB 443 GSRHGATQMEVNVWSEAIRTPAQVGFQCPHNDFEASRAGP--FAAPKVPADVTQGVNDREA 501  
OY 300 TWTGASKYNLNGRESITNPGTAMASHKDEDKF-----PMSGVMTFGKESAGASN 350  
DB 502 NGSVRSYSGKHGEMNAHGPAPERVTWDETTFGSGRDRDGFIOGAPLVVPPPLNGILT 561  
OY 351 TALDNVMTDEEIKATNPVATERFGTVAVNFQS--SSTDPAATGVHAGALPGMWQDR 408  
DB 562 NA-----NPIGTKN---DIHFSNVFNSYGPLTAFSH-----PS 591  
OY 409 DVLVQGPWAK--IPHTDGHFHPSPLMGFGGLKNPPQILIKNTP--VPANPPAEFSAT 463  
DB 592 PVYPQGIWQKELDLHKRPLATAPFV---CKNNAFGQMLVRLGPLNLTDQYDENGATLS 648  
OY 464 KFAFITQYSTGQVSEIEMELQKENSKRWNPEVOYTSNYAKSANVDFTVNNNG 517  
DB 649 RIVTYGTFFWKGLTMRAKLRA---NTTWNPNVYQ-----VSVEDNG 686

RESULT 12  
COAT\_MEVA STANDARD; PRT; 722 AA.  
ID COAT\_MEVA  
AC P27437;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Coat protein VP1 [Contains: Coat protein VP2].  
OS Mink enteritis virus (strain Abashiri) (MEV).  
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.  
OX NCBI\_Taxid=10793;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91202123; PubMed=2016597;  
RA Kariatsumari T., Horiuchi M., Hama E., Yaguchi K., Ishiguro N.,  
RA Goto H., Shinagawa M.;  
RT "Construction and nucleotide sequence analysis of an infectious DNA  
RT clone of the autonomous parvovirus, mink enteritis virus.";  
RL J. Gen. Virol. 72:867-875(1991).  
CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; D00765; BAA00663.1; -.  
DR PIR; B38350; VCPVME.  
DR HSSP; P30129; ADPV.  
DR InterPro; IPR001403; Parvo\_coat.  
DR Pfam; PF00740; Parvo\_coat; 1.  
KW Coat protein; Glycoprotein.  
FT CHAIN 1 722 COAT PROTEIN VP1.  
FT CHAIN 139 722 COAT PROTEIN VP2.  
FT DOMAIN 160 177 GLY-RICH.  
FT CARBOHYD 163 163 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 185 163 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 202 202 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 318 318 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 581 581 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 643 643 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 655 655 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 722 AA; 79823 MW; 9DADBCAB2EF9F622 CRC64;

Query Match 6.7%; Score 194; DB 1; Length 722;  
Best Local Similarity 20.7%; Pred. No. 2.1e-06;  
Matches 136; Conservative 73; Mismatches 215; Indels 232; Gaps 32;

OY 5 GGAPMADNNEGADGVNAGNWHCDSTWLGDRVITST-----RTWALPTNN- 52  
DB 148 GGQP-AVNRERATGSGNGSGGGGGS--GGVGISTGTFFNNQTEFKFLENGWEITANGS 204  
OY 53 ---HL-----YKQI-----SSASTGASNDNHYFGYSTPWGYFDENFRHCHFSFRDW 95  
DB 205 RLVHLMNESENYKRVVNVNMDKTAVKGNMALDTHVQIVTWPMSLVDANAMGWVFNPGDW 264  
OY 96 ORLNNMGFRPKRLNFKLFNIQKEVT---TNDGVTTIANNLSTVQVPSDSEYQLPYV 152  
DB 265 QLIWNTMSELHLVSEFOEIFNVVLKTVSESATQPTKYVNNNDLTASIMVALDSNNTMPTF 324  
OY 153 LGSAHQGL-----PPFPA-----DVFMIPQYGYLTNNNGSAVGRS----- 189  
DB 325 PAAMRSETLGFYFWKPTIPTPWRYFFQWDRTLIPSH-----TGTSGTPTNIYHGTDPDD 378  
OY 190 -SFYCLE-YFPQSMLRTGNF---TF-----SYTFEE-----VPFHSSYASQSL 229  
DB 379 VQFYTIENSVPVHLRTGDEFATGTFPFDCPKRLTHTWQTNRALGLPPLNSLPQSEGA 438  
OY 230 DR-----LMNPLIDQYL--YYLNRTONOS----- 251

Db 439 TNFGDIGVQDQKRGVTVQMGNTDYITEATIMRPAEYGSAPYYSFEASTQGPFKPIAAG 498  
 QY 252 -GSAQNKDLFSRGSFAGMSVQPKNMLPGPCY---RQQRVSKTKTDNNNSNFTWTGASKY 307  
 Db 499 RGAQTDENQADGDPF-----YAFGRQHGQKTTTGTETPERFTY----- 538  
 QY 308 NLNGRESIINPGTAMASHKDEDEKFFPMGSMVIFGKESAGASNTALD-NVMITDEEIK 366  
 Db 539 -----IAHQDT-----GRYPAGDWIQININFLPTNDNVLLP 570  
 QY 367 TNPVATERFGTVAVNFQSSSTDPATGDVHAMGALPGMWQDRDVLQGPWAKIPHTDGH 426  
 Db 571 TDPFG-----GKTGINY--TNIFNTYGPLTALNNVP-----PVYPNGQIMDKFEDTD-- 615  
 QY 427 FHPSPLMGFGGLK-----NPPQILIKNTPVAN---PRAEFSATKFAFIT 470  
 Db 616 -----LKPRLHVNAPFVCCNNCPGQLFVKVAPNLITNEYDPDASANMSR-----IV 660  
 QY 471 QYSTGVSEIEMELQKENSKRMPREVQYTSNYAKSANVDFTVNNGLYTEPRPIG 526  
 Db 661 TYSDFWWMKGLVFKAKLRASHTWNPICQMSIN-----VDNQFNYL-PNNIG 705

RESULT 13

COAT FPV STANDARD; PRT; 727 AA.

AC P04864; Q65112;  
 DT 13-AUG-1987 (Rel. 05, Created)  
 DT 13-AUG-1987 (Rel. 05, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Coat protein VP1 (Contains: Coat protein VP2).  
 OS Feline panleukopenia virus (FPV).  
 OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.  
 OX NCBI\_TaxID=10786;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=85265017; PubMed=2991581;  
 RA Carlson J., Rushlow R., Maxwell I., Maxwell F., Winston S., Hahn W.;  
 RT "Cloning and sequence of DNA encoding structural proteins of the  
 autonomous parvovirus feline panleukopenia virus.";  
 RL J. Virol. 55:574-587(1985).  
 CC -|- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 or send an email to license@isb-sib.ch).  
 CC -----

DR EMBL; M10824; AAA47161.1; -  
 DR EMBL; M10824; AAA47162.1; -  
 DR PIR; A03701; VCPVIF.  
 DR HSSP; P30129; 4DPV.  
 DR InterPro; IPR001403; Parvo\_coat.  
 DR Pfam; PF00740; Parvo\_coat; 1.  
 KW Coat protein; Glycoprotein.  
 RN CHAIN 1 727  
 FT CHAIN 144 727 COAT PROTEIN VP1.  
 FT CARBOHYD 168 727 COAT PROTEIN VP2.  
 FT CARBOHYD 190 168 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 190 190 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 207 207 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 323 323 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 323 323 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 586 586 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 648 648 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 660 660 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT DOMAIN 165 180 GLY-RICH.  
 SQ SEQUENCE 727 AA; 80344 MW; 77E0F6FE554C0C6E CRC64;

Query Match 6.7%; Score 194; DB 1; Length 727;  
 Best Local Similarity 20.7%; Pred. No. 2.1e-06;

Matches 136; Conservative 73; Mismatches 215; Indels 232; Gaps 32;  
 QY 5 GGAPMADNNEGADGVGNAGNWHCDSTWLGDRTVITST-----RTWALPTYNN- 52  
 Db 153 GGQP-AVRNERATGSGNGSGGGGGS--GGVGISTGFENNQTEFKLENGWEITANSS 209  
 QY 53 ---HL-----YKQI-----SSASTGASNDNHFGYSTPMWGYFDENRFCHFSPPDW 95  
 Db 210 RLVLNMPESSENYKRVVNVNMDKTAVKGNMALLDTHVQITPWSLVDANAWGWENPGDW 269  
 QY 96 QRLNNNWGFRPKRLNFKLNIQYKEVT--TNDGVTIANNLSTVQVFSDEYQLPYV 152  
 Db 270 QLIYNTMSELHLVSFEQEIFNVVLKTVSESATQPTKYVNNDLTASLMVALDSNNTMPTT 329  
 QY 153 LGSAGQCL-----PPFPA-----DVFMIPQGYTLNNGSQAVGRS----- 189  
 Db 330 PAMRSETLGFYPWKPTIPTWRYFQWDRTLIPSH-----TGTSGPTNIYHGTDPDD 383  
 QY 190 -SFYCLE-YFPSQMLRTGNF--TF-----SYTPEE-----VPFHSYAHSGSL 229  
 Db 384 VQFYTENSVPVHLRTGTDEFATGTFEFDCKPCRLTHWQNRALGLPPLNSLPQSEGA 443  
 QY 230 DR-----LMNPLIDQYL--YYLNRTQNS----- 251  
 Db 444 TNFGDIGVQDQKRGVTVQMGNTDYITEATIMRPAEYGSAPYYSFEASTQGPFKPIAAG 503  
 QY 252 -GSAQNKDLFSRGSFAGMSVQPKNMLPGPCY---RQQRVSKTKTDNNNSNFTWTGASKY 307  
 Db 504 RGAQTDENQADGDPF-----YAFGRQHGQKTTTGTETPERFTY----- 543  
 QY 308 NLNGRESIINPGTAMASHKDEDEKFFPMGSMVIFGKESAGASNTALD-NVMITDEEIK 366  
 Db 544 -----IAHQDT-----GRYPAGDWIQININFLPTNDNVLLP 575  
 QY 367 TNPVATERFGTVAVNFQSSSTDPATGDVHAMGALPGMWQDRDVLQGPWAKIPHTDGH 426  
 Db 576 TDPFG-----GKTGINY--TNIFNTYGPLTALNNVP-----PVYPNGQIMDKFEDTD-- 620  
 QY 427 FHPSPLMGFGGLK-----NPPQILIKNTPVAN---PRAEFSATKFAFIT 470  
 Db 621 -----LKPRLHVNAPFVCCNNCPGQLFVKVAPNLITNEYDPDASANMSR-----IV 665  
 QY 471 QYSTGVSEIEMELQKENSKRMPREVQYTSNYAKSANVDFTVNNGLYTEPRPIG 526  
 Db 666 TYSDFWWMKGLVFKAKLRASHTWNPICQMSIN-----VDNQFNYL-PNNIG 710

RESULT 14

COAT PAVC2 STANDARD; PRT; 584 AA.

AC P30129;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE Coat protein VP2.  
 OS Canine parvovirus (type 2 / strain A72) (CPV).  
 OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.  
 OX NCBI\_TaxID=31597;  
 RN [1]  
 RP X-RAY CRYSTALLOGRAPHY (3.25 ANGSTROMS).  
 RX MEDLINE=91173306; PubMed=2006420;  
 RA Tsao J., Chapman M.S., Agbandje M., Keller W., Smith K., Wu H.,  
 RA Luo M., Smith T.J., Rossmann M.G., Compans R.W., Parish C.R.;  
 RT "The three-dimensional structure of canine parvovirus and its  
 functional implications.";  
 RL Science 251:1456-1464(1991).  
 CC -|- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS,  
 OF A COMBINATION OF VP2, VP3, AND SOME VP1.  
 CC -|- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.  
 CC -----  
 DR PDB; 4DPV; 01-APR-97.  
 DR PDB; 1IJS; 23-DEC-96.  
 DR InterPro; IPR001403; Parvo\_coat.  
 DR Pfam; PF00740; Parvo\_coat; 1.

KW	Coat protein; Glycoprotein; 3D-structure.	
FT	CARBOHYD	25
FT	CARBOHYD	47
FT	CARBOHYD	64
FT	CARBOHYD	180
FT	CARBOHYD	443
FT	CARBOHYD	505
FT	CARBOHYD	517
FT	CARBOHYD	38
FT	STRAND	49
FT	STRAND	58
FT	STRAND	67
FT	STRAND	80
FT	STRAND	87
FT	STRAND	102
FT	STRAND	118
FT	STRAND	123
FT	STRAND	133
FT	STRAND	164
FT	STRAND	169
FT	STRAND	173
FT	STRAND	179
FT	STRAND	188
FT	STRAND	199
FT	STRAND	203
FT	STRAND	206
FT	STRAND	215
FT	STRAND	231
FT	STRAND	238
FT	STRAND	246
FT	STRAND	252
FT	STRAND	261
FT	STRAND	266
FT	STRAND	270
FT	STRAND	273
FT	STRAND	300
FT	STRAND	310
FT	STRAND	317
FT	STRAND	327
FT	STRAND	330
FT	STRAND	335
FT	STRAND	339
FT	STRAND	346
FT	STRAND	352
FT	STRAND	363
FT	STRAND	367
FT	STRAND	373
FT	STRAND	378
FT	STRAND	388
FT	STRAND	396
FT	STRAND	410
FT	STRAND	415
FT	STRAND	418
FT	STRAND	426
FT	STRAND	429
FT	STRAND	445
FT	STRAND	450
FT	STRAND	453
FT	STRAND	455
FT	STRAND	464
FT	STRAND	465
FT	STRAND	470
FT	STRAND	488
FT	STRAND	490
FT	STRAND	498
FT	STRAND	506
FT	STRAND	512
FT	STRAND	520
FT	STRAND	523
FT	STRAND	555
FT	STRAND	558
FT	STRAND	565
FT	STRAND	574

SQ	SEQUENCE	584 AA;	64686 MW;	6E4DADA5AEBF9D8C CRC64;
QY	Query Match	6.5%;	Score 190;	DB 1; Length 584;
Db	Best Local Similarity	20.3%;	Pred. No. 3e-06;	
QY	Matches 130; Conservative	82;	Mismatches 229;	Indels 198; Gaps 30;
QY	5	GGAPMADNNEGADGVNASGNWHCDSTWLGDRIITST-----RTWALPTYNN-	52	
QY	10	GGQP-AVNERATGSGNGSGGGGGGS--GGVGISTGTENNQTETKFLENGWVYITANGS	66	
QY	53	---HL-----YKQI-----SSASTGASNDNHYFGYSTPGYFDENRFCHFSPRDW	95	
QY	67	RLVHLNMPSESENYRRVNVNMMDKTAVNGNMALDDIHAETVTPMSLVDANAMGWFMNPGDW	126	
QY	96	QRLINNNGFRPPKRLNFKLFNIQYKEVT--TNDGVTIANNLSTVQVFSDEYQLPYV	152	
QY	127	QLIVNTMSEHLVSPFQEIFNVVLKTVSESATQPTKVYNNDLTASLWALDSNNTMPFT	186	
QY	153	LGSAGGCL-----PPFPA-----DVFMIPOGYLTLLNGSQAVGRS-----	189	
QY	187	PAAMRSETLGFYPMWKPTIPTPMRYFYQWDRTLIPSH-----TGTSGTPTNIYHGTDPD	240	
QY	190	-SFYCLE-YFPSQMLRTGNF-TFSYTFEEVPEFHSYAHSGSLDRLMN--PLIDQLYYL	244	
QY	241	VQFYTIENSVFVHLRLTGDEFATGTFEFDCKP--CRLTHTWQTNALGLP-----FL	291	
QY	245	NRTQNGSGAQNKDLFSRGSFAGMSVQPKNWLPGPCYRQQRVSKTKTDNNNSNF-----	299	
QY	292	NSLPQSEGAATNFGDI-----GV-----QODRARGVTQMGTNTYITEAT	329	
QY	300	---TWGASKYNLNGRESTINP-----GTAMASHKDEDEKFFEMSGVMIFGKESAGASN	350	
QY	330	IMRPAEVGYSAFYSPFASTQGPFTPIAAGRGGAQTDENQADGNPRYAFGRQHGKKT	389	
QY	351	TALD-----NMTDEBEIKATNPVATERFGTVAVNFQ	383	
QY	390	TTGETPERFTYIAHQDTGRYPEDMIQININFLPTNDNVLLPTDPDG---GKTGINY-	444	
QY	384	SSSTDPAAGDVHAMGALPGMWODRDVYLQGPWAKIPIHTDGHFHPSPLMGGFGLK----	439	
QY	445	-TNIFNTYGPLTALNNVP-----VYPNGQIMWKEFDTD-----LKPRHL	483	
QY	440	-----NPPQILIKNTPVPAN--PFAEFSATKFAFPIQYSTGQVSVEIEWELQK	487	
QY	484	VNAPFVCONNCPGLFVKVAPNLITQYDIDASANMSR---IVTYSDFWWMKGKLVFXAKL	539	
QY	488	ENSKRMNPEVQYTSNYAKSANVDFTVNNGLYTEPRPIG	526	
QY	540	RASHTMNPICQMSIN-----VDNQFNYY-PSNIG	567	
RC	RESULT 15			
RC	COAT_FPV19	STANDARD;	PRT;	727 AA.
RC	ID	COAT_FPV19		
RC	AC	P24840;		
RC	DT	01-MAR-1992 (Rel. 21, Created)		
RC	DT	01-MAR-1992 (Rel. 21, Last sequence update)		
RC	DT	16-OCT-2001 (Rel. 40, Last annotation update)		
RC	DE	Coat protein VP1 (Contains: Coat protein VP2).		
RC	OS	Feline panleukopenia virus (strain 193) (FPV).		
RC	OC	Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.		
RC	OX	NCBI_TaxID=10787;		
RC	RN	[1]		
RC	RP	SEQUENCE FROM N.A.		
RC	RC	STRAIN=193/70;		
RC	RX	MEDLINE=91073139; PubMed=2174965;		
RC	RA	Martyn J.C., Davidson B.E., Studdert M.J.;		
RC	RT	"Nucleotide sequence of feline panleukopenia virus: comparison with		
RC	RT	canine parvovirus identifies host-specific differences.";		
RC	RL	J. Gen. Virol. 71:2747-2753(1990).		
RC	RP	[2]		
RC	RC	SEQUENCE FROM N.A.		
RC	RC	STRAIN=CU-4;		



RX MEDLINE=91272479; PubMed=1647068;  
RA Parrish C.R.;  
RT "Mapping specific functions in the capsid structure of canine  
RT parvovirus and feline panleukopenia virus using infectious plasmid  
RT clones";  
RL Virology 183:195-205(1991).  
CC -|- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X55115; CAA38911.1; -  
DR EMBL; M38246; AAC37928.1; -  
DR EMBL; M38246; AAC37929.1; -  
DR PIR; B36608; VCPYFP.  
DR HSSP; P30129; 4BPV.  
DR InterPro; IPR001403; Parvo\_coat.  
DR Pfam; PF00740; Parvo\_coat; 1.  
KW Coat protein; Glycoprotein.  
FT CHAIN 1 727 COAT PROTEIN VP1.  
FT CHAIN 1 727 COAT PROTEIN VP2.  
FT DOMAIN 144 727 GLY-RICH.  
FT CARBOHYD 165 190 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 168 190 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 190 190 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 207 207 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 323 323 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 586 586 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 648 648 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 660 660 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT DOMAIN 165 180 GLY-RICH.  
SQ SEQUENCE 727 AA; 80386 MW; 648596C09B621FF5 CRC64;

Query Match 6.5%; Score 189; DB 1; Length 727;  
Best Local Similarity 20.6%; Pred. No. 4.8e-06;  
Matches 135; Conservative 73; Mismatches 216; Indels 232; Gaps 32;

QY 5 GGAPMADNNEGADGVGNASGNWCHDSTWLGDRVITST-----RTWALPTYNN- 52  
DB 153 GGQP-AVRNERATGSGNGSGGGGGG--GGVGISTGTENNQTEFKPLENGWEITANSS 209  
QY 53 ---HL-----YKQI-----SSASTGASNDNHYFGYSTPWGYFDENRFHCHFSPRDW 95  
DB 210 RLVLHLMPESENYKRVVNNMMDKTAVKGNALDDIHVQIVTPMSLV DANAMGVWFNPGDW 269  
QY 96 QRLINNNMGFRPKRLNFKLFNIQVKEVT---TNDGVTIANNLSTVQVFSDEYQLPYV 152  
DB 270 QLIYNTMSELHLVSFEQEIFNVVLKTVSESATQPTKVYNNNDLTASLMVALDSNNTMPT 329  
QY 153 LGSAGGCL-----PPPA-----DVFMIPOGYLTLNNGSAVGRS----- 189  
DB 330 PAAMRSEITGFYPWKPTIPTPWRYTFQWDRLLIPSH-----TGTSPTNVYHGTDPDD 383  
QY 190 -SFYCLE-YFPSQMLRTGNF--TF-----SYFEE-----VPFHSSYAHSQSL 229  
DB 384 VQFYTIENGVVHLRLRTGDEFATGTFEFDCKPCLRLTHWTQNRALGLPPFLNSLPQSEGA 443  
QY 230 DR-----LMNPLIDQYL--YYLNRTQNS----- 251  
DB 444 TNFGDIGVQODKRRGVTOMGNTDYITTEATIMRPAEVGYSAPYSPFEASTQGPKTPIAAG 503  
QY 252 -GSAQNKDLLFSRGSFAGMSVQPKNWLPGPCY--RQQRVSKTKTDNNSNFTWTGASKY 307  
DB 504 RGAQGTDEQAADGDP-----YAFGRQHGQKTTTGETPERFTY----- 543  
QY 308 NLNGRESINPGTAMASHKDEDEKFFPMGVMIFGKESAGASNTALD-NVMITDEEIK 366  
DB 544 -----IAHODT-----GRYPEGDWIONINFNLPVTNDNVLLP 575

QY 367 TNPVATERFGTVAVNFQSSSTDPAIGDVHAMGALPGWVQDRDVLQGPWAKIPHTDGH 426  
DB 576 TDPIC---GKTGINY--TNIFNTYGLTALNNVP-----PVYPNGQIWDKEPDTD-- 620  
QY 427 FHPSPLMGGFLK-----NPPQILIKTPVEAN---PPAEFSATKFAFIT 470  
DB 621 -----LKPRLHVNAPFVCQNNCPGQLFVKVAPNLTNEYDPDASANMSR-----IV 665  
QY 471 QYSTQVSEIEMWELQKENSRRWNPBEVQYTSNYAKSANDVFTVDNNGLYTEPRPIG 526  
DB 666 TVSDFWWKGKLVFKAKLRASHTWNPQQMSIN-----VDNQFNTV-PNNIG 710

Search completed: July 17, 2003, 18:32:07  
Job time : 12.5714 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 17, 2003, 18:26:09 ; Search time 39.7143 Seconds  
(without alignment)  
2770.518 Million cell updates/sec

Title: US-09-807-802A-17  
Perfect score: 2906  
Sequence: 1 MASGGAPMADNNEGADGVG.....NNGLYTEPRPIGTRYLTRPL 534

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL 21:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_rvirus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2906	100.0	736	12 Q9WBP8	Q9wbp8 adeno-associ
2	2884	99.2	736	12 O56137	O56137 adeno-associ
3	2530	87.1	736	12 O56139	O56139 adeno-associ
4	2528	87.0	736	12 Q65311	Q65311 adeno-associ
5	2481.5	85.4	533	12 O92917	O92917 adeno-associ
6	2481.5	85.4	598	12 O56653	O56653 adeno-associ
7	2481.5	85.4	735	12 O56652	O56652 adeno-associ
8	1698	58.4	534	12 Q67668	Q67668 goose parvo
9	1698	58.4	587	12 Q67667	Q67667 goose parvo
10	1698	58.4	732	12 Q67666	Q67666 goose parvo
11	1697	58.4	732	12 Q8V395	Q8v395 goose parvo
12	1690.5	58.2	734	12 O41855	O41855 adeno-associ
13	1690	58.2	732	12 O83290	O83290 muscovy duc
14	1676	57.7	534	12 Q65446	Q65446 barbarie du
15	1676	57.7	587	12 Q65445	Q65445 barbarie du
16	1676	57.7	732	12 Q65444	Q65444 barbarie du

17	1676	57.7	732	12 Q83289	Q83289 muscovy duc
18	1665	57.3	724	12 Q9Y1J1	Q9y1j1 adeno-associ
19	1521	52.3	676	12 Q67672	Q67672 goose parvo
20	715	24.6	179	12 Q9WA24	Q9wa24 goose parvo
21	715	24.6	179	12 Q9W8U7	Q9w8u7 duck parvo
22	711	24.5	179	12 Q9WN18	Q9wn18 duck parvo
23	709	24.4	179	12 Q9WN19	Q9wn19 duck parvo
24	707	24.3	179	12 Q9WN20	Q9wn20 goose parvo
25	614	21.1	571	12 Q8Q0Y4	Q8q0y4 minute viru
26	614	21.1	703	12 Q8Q0Y5	Q8q0y5 minute viru
27	536.5	18.5	947	12 Q918U9	Q918u9 bovine parv
28	501	17.2	554	12 Q9PZS9	Q9pzs9 human parvo
29	501	17.2	781	12 Q9PZT0	Q9pzt0 human parvo
30	495	17.0	785	12 Q9J0X4	Q9j0x4 pig-tailed
31	494.5	17.0	554	12 Q9JGP7	Q9jgp7 human parvo
32	494.5	17.0	781	12 Q9JGP8	Q9jgp8 human parvo
33	493	17.0	546	12 Q913X0	Q913x0 human parvo
34	493	17.0	773	12 Q913X1	Q913x1 human parvo
35	492.5	16.9	781	12 P89317	P89317 human parvo
36	492.5	16.9	781	12 P89318	P89318 human parvo
37	492	16.9	781	12 P89319	P89319 human parvo
38	490.5	16.9	781	12 P89316	P89316 human parvo
39	490	16.9	554	12 Q912B7	Q912b7 human eryth
40	490	16.9	554	12 Q90201	Q90201 human parvo
41	490	16.9	769	12 Q9PZT4	Q9pzt4 human parvo
42	490	16.9	781	12 Q9JGS0	Q9jgs0 human parvo
43	490	16.9	781	12 Q912B8	Q912b8 human eryth
44	490	16.9	781	12 P90221	P90221 human parvo
45	490	16.9	781	12 P90222	P90222 human parvo

ALIGNMENTS

RESULT 1									
ID	Q9WBP8	PRELIMINARY;	PRT;	736	AA.				
AC	Q9WBP8;								
DT	01-NOV-1999 (TREMBlrel. 12, Created)								
DT	01-NOV-1999 (TREMBlrel. 12, Last sequence update)								
DT	01-JUN-2001 (TREMBlrel. 17, Last annotation update)								
DE	Capsid protein.								
OS	adeno-associated virus 1.								
OC	Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.								
OX	NCBI_TaxID=85106;								
RN	[1]								
RP	SEQUENCE FROM N.A.								
RX	MEDLINE=99214338; PubMed=10196295;								
RA	Xiao W., Chirmule N., Berta S.C., McCullough B., Gao G., Wilson J.M.;								
RT	"Gene therapy vectors based on adeno-associated virus type 1.";								
RL	J. Virol. 73:3994-4003(1999).								
RN	[2]								
RP	SEQUENCE FROM N.A.								
RA	Xiao W., Wilson J.M.;								
RL	Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.								
DR	EMBL; AF063497; AAD27757.1; -.								
DR	InterPro; IPR001403; Parvo_coat.								
DR	Pfam; PF00740; Parvo_coat; 1.								
SQ	SEQUENCE 736 AA; 81375 MW; CFAFB9BD5CD0595 CRC64;								
Query Match									
Best Local Similarity 100.0%; Score 2906; DB 12; Length 736;									
Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0									
QY	1	MASGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVITSTRTVALPTYNHLYKQISS	60						
DB	203	MASGGAPMADNNEGADGVGNAGNWHCDSTWLGDRVITSTRTVALPTYNHLYKQISS	262						
QY	61	ASTGASNDNHYPFGYSTPWGYPFDRNFHCHFSPRDWQRLINNNWGFRPKRLNFKLFNIQVK	120						
DB	263	ASTGASNDNHYPFGYSTPWGYPFDRNFHCHFSPRDWQRLINNNWGFRPKRLNFKLFNIQVK	322						
QY	121	EVTNDGVTTIANNLSTVQVESDSEYQLPYVLGSAHQGCLPPFPADVFMIPQYGYLTIN	180						



|||||  
Db 323 EYTTNDGVTTIANNLSTVQVFSDEYQLPYLGSAHQGLPPFPADVEMIPQYGYLTIN 382  
Qy 181 NGSQAVGRSSFYCLEYFPSPQMLRTGNNFTFSYTFEEVPHSSSYAHSQSLDRLMPLIDY 240  
Db 383 NGSQAVGRSSFYCLEYFPSPQMLRTGNNFTFSYTFEEVPHSSSYAHSQSLDRLMPLIDY 442  
Qy 241 LYYLNRTONQSGSAQNKDLLFSRGS PAGMSVOPKNWLPGPCYRQQRVSKTKTDNNNSNFT 300  
Db 443 LYYLNRTONQSGSAQNKDLLFSRGS PAGMSVOPKNWLPGPCYRQQRVSKTKTDNNNSNFT 502  
Qy 301 WTGASKYNLNGRESIINPGTAMASHKDDKDFPMSGVMI FGKESAGASNTALDNVMITD 360  
Db 503 WTGASKYNLNGRESIINPGTAMASHKDDKDFPMSGVMI FGKESAGASNTALDNVMITD 562  
Qy 361 EEEIKATNPVATERFGTVAVNFQSSSTD PATGDVHAMGALPGMWQDRDYYLQGPIMAKI 420  
Db 563 EEEIKATNPVATERFGTVAVNFQSSSTD PATGDVHAMGALPGMWQDRDYYLQGPIMAKI 622  
Qy 421 PHTDGHFHPSPPLMGFGFLKNPPQILIKNTVPANPPAEFSATKFAFITQYSTGQVSVE 480  
Db 623 PHTDGHFHPSPPLMGFGFLKNPPQILIKNTVPANPPAEFSATKFAFITQYSTGQVSVE 682  
Qy 481 IEWELQKENSKRWNPEVOYTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 534  
Db 683 IEWELQKENSKRWNPEVOYTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 736

RESULT 2  
056137  
ID 056137 PRELIMINARY; PRT; 736 AA.  
AC 056137;  
DT 01-JUN-1998 (TReMBLrel. 06, Created)  
DT 01-JUN-1998 (TReMBLrel. 06, Last sequence update)  
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
DE Capsid protein VP1.  
OS adeno-associated virus 6.  
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.  
OX NCBI\_TaxID=68558;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98080418; PubMed=9420229;  
RA Rutledge E.A., Halbert C.L., Russell D.W.;  
RT "Infectious clones and vectors derived from adeno-associated virus  
RT (AAV) serotypes other than AAV type 2.";  
RL J. Virol. 72:309-319(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Rutledge E.A., Russell D.W.;  
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF028704; AAB95450.1;  
DR InterPro; IPR001403; Parvo\_coat.  
DR Pfam; PF00740; Parvo\_coat; 1.  
SQ SEQUENCE 736 AA; 81411 MW; 311217A089C565F5 CRC64;

Query Match 99.2%; Score 2884; DB 12; Length 736;  
Best Local Similarity 99.1%; Pred. No. 3.1e-204;  
Matches 529; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MASGGGAPMADNNEGADGVGNASGNWHCDSTWLGDRIITSTRTVALPTYNNHLYKQISS 60  
Db 203 MASGGGAPMADNNEGADGVGNASGNWHCDSTWLGDRIITSTRTVALPTYNNHLYKQISS 262  
Qy 61 ASTGASNDNHYFGYSTPWGYFDENRFHCHFSPRDWQRLINNNGWFRPKRLNFKLFNIQVK 120  
Db 263 ASTGASNDNHYFGYSTPWGYFDENRFHCHFSPRDWQRLINNNGWFRPKRLNFKLFNIQVK 322  
Qy 121 EYTTNDGVTTIANNLSTVQVFSDEYQLPYLGSAHQGLPPFPADVEMIPQYGYLTIN 180  
Db 323 EYTTNDGVTTIANNLSTVQVFSDEYQLPYLGSAHQGLPPFPADVEMIPQYGYLTIN 382  
Qy 181 NGSQAVGRSSFYCLEYFPSPQMLRTGNNFTFSYTFEEVPHSSSYAHSQSLDRLMPLIDY 240

Db 383 NGSQAVGRSSFYCLEYFPSPQMLRTGNNFTFSYTFEDVPFHSSSYAHSQSLDRLMPLIDY 442  
Qy 241 LYYLNRTONQSGSAQNKDLLFSRGS PAGMSVOPKNWLPGPCYRQQRVSKTKTDNNNSNFT 300  
Db 443 LYYLNRTONQSGSAQNKDLLFSRGS PAGMSVOPKNWLPGPCYRQQRVSKTKTDNNNSNFT 502  
Qy 301 WTGASKYNLNGRESIINPGTAMASHKDDKDFPMSGVMI FGKESAGASNTALDNVMITD 360  
Db 503 WTGASKYNLNGRESIINPGTAMASHKDDKDFPMSGVMI FGKESAGASNTALDNVMITD 562  
Qy 361 EEEIKATNPVATERFGTVAVNFQSSSTD PATGDVHAMGALPGMWQDRDYYLQGPIMAKI 420  
Db 563 EEEIKATNPVATERFGTVAVNFQSSSTD PATGDVHAMGALPGMWQDRDYYLQGPIMAKI 622  
Qy 421 PHTDGHFHPSPPLMGFGFLKNPPQILIKNTVPANPPAEFSATKFAFITQYSTGQVSVE 480  
Db 623 PHTDGHFHPSPPLMGFGFLKNPPQILIKNTVPANPPAEFSATKFAFITQYSTGQVSVE 682  
Qy 481 IEWELQKENSKRWNPEVOYTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 534  
Db 683 IEWELQKENSKRWNPEVOYTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRPL 736

RESULT 3  
056139  
ID 056139 PRELIMINARY; PRT; 736 AA.  
AC 056139;  
DT 01-JUN-1998 (TReMBLrel. 06, Created)  
DT 01-JUN-1998 (TReMBLrel. 06, Last sequence update)  
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
DE Capsid protein VP1.  
OS adeno-associated virus 3B.  
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.  
OX NCBI\_TaxID=68742;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98080418; PubMed=9420229;  
RA Rutledge E.A., Halbert C.L., Russell D.W.;  
RT "Infectious clones and vectors derived from adeno-associated virus  
RT (AAV) serotypes other than AAV type 2.";  
RL J. Virol. 72:309-319(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Rutledge E.A., Russell D.W.;  
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF028705; AAB95452.1;  
DR InterPro; IPR001403; Parvo\_coat.  
DR Pfam; PF00740; Parvo\_coat; 1.  
SQ SEQUENCE 736 AA; 81906 MW; DD52331AD5F0D70F CRC64;

Query Match 87.1%; Score 2530; DB 12; Length 736;  
Best Local Similarity 86.0%; Pred. No. 3.9e-178;  
Matches 460; Conservative 27; Mismatches 46; Indels 2; Gaps 2;

Qy 1 MASGGGAPMADNNEGADGVGNASGNWHCDSTWLGDRIITSTRTVALPTYNNHLYKQISS 60  
Db 203 MASGGGAPMADNNEGADGVGNSSGNWHCDSQWLGDRIITSTRTVALPTYNNHLYKQISS 262  
Qy 61 ASTGASNDNHYFGYSTPWGYFDENRFHCHFSPRDWQRLINNNGWFRPKRLNFKLFNIQVK 120  
Db 263 QS-GASNDNHYFGYSTPWGYFDENRFHCHFSPRDWQRLINNNGWFRPKLSFKLFNIQVK 321  
Qy 121 EYTTNDGVTTIANNLSTVQVFSDEYQLPYLGSAHQGLPPFPADVEMIPQYGYLTIN 180  
Db 322 EYTTNDGVTTIANNLSTVQVFTDSEYQLPYLGSAHQGLPPFPADVEMIPQYGYLTIN 381  
Qy 181 NGSQAVGRSSFYCLEYFPSPQMLRTGNNFTFSYTFEEVPHSSSYAHSQSLDRLMPLIDY 240  
Db 382 NGSQAVGRSSFYCLEYFPSPQMLRTGNNFQFSYTFEDVPFHSSSYAHSQSLDRLMPLIDY 441  
Qy 241 LYYLNRTO-NQSGSAQNKDLLFSRGS PAGMSVOPKNWLPGPCYRQQRVSKTKTDNNNSNFT 299  
Db 442 LYYLNRTOGTTSGTTNQSRLLFQAGFQSMQLQARNWLPGPCYRQQRVSKTANDNNNSNFT 501

OY	300	TWTGASKYNLNGRESIIINPGTAMASHKDEDEDKEEPMSCVMI FGKESAGASANTALDNVMIT	359
Db	502	PMTAASKYHLNGRDSLVPNGPAMASHKODEEXFEPMHGNIIFGEGETTASNAEIDNMWIT	561
OY	360	DEEBIKATNPVATERFGTVAVNFQSSSTDPAZGDVHAMGALPGMWODRDVYLQGPIMAK	419
Db	562	DEEIRRTNPVATEQYGTVANNLQSSNTAPTTRTVNDQGALLPGMVWQDRDVLQGPIMAK	621
OY	420	IPHTDGHFHPSPILMGFGFLKNPPQILIKNTPYBPANPNPAEFSA TKFASFITQYSTGOVS	479
Db	622	IPHTDGHFHPSPLMGFGFLKHPPQIMIKNTPYBPANPPTTFSPA KFAFSFITQYSTGOVS	681
OY	480	EIEMELOKENSKRWNPEVOYTSNYAKSANDFTVDNNGLYTEPRPIGTRYLTRPL	534
Db	682	EIEMELOKENSKRWNPEIQYTSNYNKSVANDFTVDNKGVSSEPRPIGTRYLTRNL	736

```

RESULT 4
Q65311
ID Q65311 PRELIMINARY; PRT; 736 AA.
AC Q65311;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE Capsid protein.
OS adeno-associated virus 3.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
OX NCBI_TaxID=46350;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3H;
RX MEDLINE=96266430; PubMed=8661429;
RA Muramatsu S., Mizukami H., Young N.S., Brown K.E.;
RT "Nucleotide sequencing and generation of an infectious clone of adeno-
RL virology 221:208-217(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=3H;
RA Muramatsu S., Brown K.E.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U48704; AAC55049.1; -.
DR InterPro; IPR001403; Parvo_coat.
DR Pfam; PF00740; Parvo_coat; 1.
SQ SEQUENCE 736 AA; 81660 MW; AFF1BF47B5C67A10 CRC64;

```

Query Match	87.0%	Score 2528;	DB 12;	Length 736;
Best Local Similarity	85.8%	Pred. No. 5.5e-178;		
Matches 459; Conservative	28;	Mismatches 46;	Indels 2;	Gaps 2;

QY	1	MASGGGAPMADNNEGADGVGNASGNWCHDSIWLGDRIITTTSTRWALPTYNHLYKQISS	60
Db	203	MASGGAPMADNNEGADGVGNSSGNWCHDSQWLGDRVITTTSTRWALPTYNHLYKQISS	262
QY	61	ASTGASNDNHYFGYSTPMGYFDFNRFHCHFSPRDWQRLINNNWGFRRPKRLNFKLFIQVK	120
Db	263	QS-GASNDNHYFGYSTPMGYFDFNRFHCHFSPRDWQRLINNNWGFRRPKLSFKLFNIQVR	321
QY	121	EYTTNDGVTTIANNLSTVQVFSDESQYLPYVLGSAHQGLPPFPADVFEMIPQYGYLTIN	180
Db	322	GYTQNDGTTIANNLSTVQVFTDSEYQLPYVLGSAHQGLPPFPADVFEMIPQYGYLTIN	381
QY	181	NGSQAVGRSSFYCLLEYFPSQMLRTGNNTFSYTFEEVPEHSSYAHQSQSLDRMLNPLIDQY	240
Db	382	NGSQAVGRSSFYCLLEYFPSQMLRTGNNFQSYTFEDVPFHSSYAHQSQSLDRMLNPLIDQY	441
QY	241	LYYLNRTQ-NQSGSAQNKDLFSRGSFAGNSVQPKNMLPGPCYRQQRVSKYTTDNNNSNF	299
Db	442	LYYLNRTQGTTSGETTNQSRLLFSQAGPQMSIQARNMLPGPCYRQQLSKTANDNNNSNF	501
QY	300	TWTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMGSMIFGKESAGASNTALDNVMT	359

Db	502	PWTAAASKYHLNGRDSLVPNGPAMASHKDDKEKFPMHGNLIFGKEGTTASNAELDNVMIT	561
QY	360	DEEEIKATNPVATERFGYVAVNFQSSSTDTPATGVDVHAMGALPGMWODRDVYLLQGPIMAK	419
Db	562	DEEIRTTNPVATEQYGYVANNLQSSNTAPTGTGVNHQALPGMWODRDVYLLQGPIMAK	621
QY	420	IPHTDGHFHPSPLMGGFGLKNPPEOILIKNTVPVPPANPPAEFSATKPAASITQYSTGQVSV	479
Db	622	IPHTDGHFHPSPLMGGFGLKHPPEOIMIKNTVPVPPANPPPTTSSPAKFAFITQYSTGQVSV	681
QY	480	EIEMELQKENSKRWNPEVOQYTSNYSKASANVDFTVDNNGLYTEPRPIGTRYLTRPL	534
Db	682	EIEMELQKENSKRWNPEIQYTSNYSKASVNVDFTVDTNGVYSEPRPIGTRYLTRNL	736

```

RESULT 5
092917
ID 092917 PRELIMINARY; PRT; 533 AA.
AC 092917;
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE Major coat protein VP3.
OS Adeno-associated virus 2 (AAV2).
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
OX NCBI_TaxID=10804;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95088582; Pubmed=7996133;
RA Ruffing M., Heid H., Kleinschmidt J.A.;
RT "Mutations in the carboxy terminus of adeno-associated virus 2 capsid
RT proteins affect viral infectivity: lack of an RGD integrin-binding
RT motif.";
RL J. Gen. Virol. 75:0-0(0).
RN [2]
RP SEQUENCE FROM N.A.
RA Berns R.I., Bohenzky R.A., Cassinotti P., Colvin D., Donahue B.A.,
RA Dull T., Horer M., Kleinschmidt J.A., Ruffing M., Snyder R.O.,
RA Tratschn J.-D., Weitz M.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF043303; AAC03779.1; -
DR InterPro; IPR001403; Parvo_coat.
DR Pfam; PF00740; Parvo_coat; 1.
KM Coat protein.
FT VARIANT 351 351 D -> N.
FT VARIANT 365 365 T -> N.
FT VARIANT 475 476 QV -> HV.
FT VARIANT 508 508 V -> R.
SQ SEQUENCE 533 AA; 60063 MW; 9E4D8BC25810D4F0 CRC64;

```

Query Match	85.4%;	Score 2481.5;	DB 12;	Length 533;
Best Local Similarity	83.3%;	Pred. No. 9.4e-175;		
Matches 445;	Conservative 37;	Mismatches 51;	Indels 1;	Gaps 1;

```

QY      1 MASGGGAPMADNNEGADGVGASGNWHCDSTWLGDRIVITTTSTRWALPTYNNHLYKQISS 60
      ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db      1 MATGSGAPMADNNEGADGVGSSGNWHCDSTWMDRIVITTTSTRWALPTYNNHLYKQISS 60
QY      61 ASTGASNDNHYFGYSTPWGYEDFNRFCHFSPRDWQRLINNNGFRPRKLNFKLFNIQVK 120
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db      61 QS-GASNDNHYFGYSTPWGYEDFNRFCHFSPRDWQRLINNNGFRPRKLNFKLFNIQVK 119
QY      121 EVTTNDGVTTIANNLTSTVQVFSDEYQLPYVLGSAHQGCLPPFPADVEMIPOGYLTLN 180
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db      120 EVTQNDGTTTIANNLSTVQVFTDSEYQLPYVLGSAHQGCLPPFPADVEMVPOGYLTLN 179
QY      181 NGSQAVGRSSFYCLEYFPSQMLRTGNNTFTSFYTFEEVPPHSSYAHQSOLDRLMNPIDQY 240
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db      180 NGSQAVGRSSFYCLEYFPSQMLRTGNNTFTSFYTFEDVPPHSSYAHQSOLDRLMNPIDQY 239
QY      241 LYYLNRTONQSGSAQNKILLFSRGS PAGMSVOPKNWLPGPCYRQORVSKTKTDNNNSNFT 300
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db      240 LYYLSRINTPSGTTTQSRIOFSQAGASDIRDOSRNWLPGPCYRQORVSKTSADNNNSEYS 299

```





QY 181 NGSQAVGRSSFYCLEYFPPSOMLRGTGNNFTSYTFEEVPHSSYAHQSOLDRLMPLIDQY 240  
DB 382 NGSQAVGRSSFYCLEYFPPSOMLRGTGNNFTSYTFEDVPFHSSYAHQSOLDRLMPLIDQY 441  
QY 241 LYYLNRTONQSGSAQNKDLLFSRGS PAGMSVQPKNMLPGPCYRQORYSKTKTDNNNSFT 300  
DB 442 LYYLSRTPSPSGTTTQSRLOFSQAGASDIRDQSRNMLPGPCYRQORYSKTSADNNSEYS 501  
QY 301 WTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMGVMIFGKESAGASNTALDNVMTD 360  
DB 502 WTGATKTHLNGRDSLVPNGPAMASHKDEKEFFQSGVLIFGKQSEKTNVDIEKMITD 561  
QY 361 EEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMGALPGMWQDRDYYLQGPIMAKI 420  
DB 562 EEEIRTNPVATEQYGSVSTNLQRGNQAATADVNTQGVLPGMWQDRDYYLQGPIMAKI 621  
QY 421 PHTDGHFHPSPLMGGFGLKNPPQILIKNTVPANPPAEFSATKFAFITQYSTGVSV 480  
DB 622 PHTDGHFHPSPLMGGFGLKHPPPQILIKNTVPANPPSTTFSAAKFAFITQYSTGVSV 681  
QY 481 IEWELQKNSKRNPEVOYTSNYAKSANVDFTVNNGLYTEPRPIGTRYLTRPL 534  
DB 682 IEWELQKNSKRNPEIOYTSNYKSVNVDFTVDTNGVYSEPRPIGTRYLTRNL 735

RESULT 8

ID Q67668 PRELIMINARY; PRT; 534 AA.  
AC Q67668;  
DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE VP3.  
GN VP3.  
OS goose parvovirus.  
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.  
OX NCBI\_Taxid=38251;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VIRULENT B;  
RX MEDLINE=96010229; PubMed=7571426;  
RA Zadori Z., Stefancsik R., Rauch T., Kisary J.;  
RT "Analysis of the complete nucleotide sequences of goose and muscovy  
RT duck parvoviruses indicates common ancestral origin with adeno-  
RT associated virus 2.";  
RL Virology 212:562-573(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VIRULENT B;  
RA Zadori Z.;  
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U25749; AAA83232.1; -.  
DR InterPro; IPR001403; Parvo\_coat.  
DR Pfam; PF00740; Parvo\_coat; 1.  
SQ SEQUENCE 534 AA; 59996 MW; F1F049558EACE92 CRC64;

Query Match 58.4%; Score 1698; DB 12; Length 534;  
Best Local Similarity 57.1%; Pred. No. 5.5e-117;  
Matches 310; Conservative 87; Mismatches 128; Indels 18; Gaps 8;

QY 1 MASGGAPMADNNEGADGVGNASGNWHCDSTWLGDRTVITSTRTWALPTYNNHLYKQISS 60  
DB 1 MAEGGGAGMGDSSSGADGVGNASGNWHCDSQMGNTVITKTRTWLPSYNNHIYKAITS 60  
QY 61 ASTGASNDNH--YFGYSTPWGYFDENRFHCHFSPRDMQRLINNNWGFRPKRLNFKLFNIQ 118  
DB 61 ---GTSQDANVOYAGYSTPWGYFDENRFHCHFSPRDMQRLINNNWGIRPKSLKFKIFNVQ 117  
QY 119 VKEVTTNDGVTTIANNLTSTVQVFSDEYQLPYVLGSAHQGLPPPADVFMIPOYGYLT 178  
DB 118 VKEVTTQDQTKTIANNLSTIQVFTDDEQDLPYVLGSAATEGTMPPEPSDVYALPOYGYCT 177  
QY 179 LN--NGSQAVGRSSFYCLEYFPPSOMLRGTGNNFTSYTFEEVPHSSYAHQSOLDRLMNP 235

DB 178 MHTNONGARFNDRSAPFYCLEYFPPSOMLRGTGNNFEFTFDFFVPHFSMAFHSQOLDRLMNP 237  
QY 236 LIDQYLYLNRTONQSGSAQNKDLLFSRGS PAGMSVQPKNMLPGPCYRQORY-SKTKTDN 294  
DB 238 LVDQYLMNFNEV-DSSRNAQ-----FKKAVKGAYGTMGKRWMLPGPKFLDQVRAYTGCTD 291  
QY 295 NNSNFT-WTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMGVMIFGKE--SAGASNT 351  
DB 292 NYAANNIMWSNGKNVNLKDRQYLLQGPVSATYTEGEASSLPAQNILGIADPYRSGSTTA 351  
QY 352 ALDNVITDEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMGALPGMWQDRDYY 411  
DB 352 GISDLMVTEBEQVAPTINGVGMKPYGRTVTNEQNTTAPTSSDLVLGALPGMWQNRDLY 411  
QY 412 LQGPIMAKIPHTDGHFHPSPLMGGFGLKNPPQILIKNTVPANPPAEFSATKFAFITQ 471  
DB 412 LQGPIMAKIPKTDGKHPSPNLGGFGLHNPQPQVFIKNTVPADPVEYVHOKMNSYITQ 471  
QY 472 YSTGVSVIEWELQKNSKRNPEVOYTSNYAKSANVDFTVNNGLYTEPRPIGTRYLT 531  
DB 472 YSTGVTEVMWELRKNSKRNPEIOYTSNFSNRTSIMPANETGCVEDRLIGTRYLT 531  
QY 532 RPL 534  
DB 532 QNL 534

RESULT 9

ID Q67667 PRELIMINARY; PRT; 587 AA.  
AC Q67667;  
DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE VP2.  
GN VP2.  
OS goose parvovirus.  
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.  
OX NCBI\_Taxid=38251;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VIRULENT B;  
RX MEDLINE=96010229; PubMed=7571426;  
RA Zadori Z., Stefancsik R., Rauch T., Kisary J.;  
RT "Analysis of the complete nucleotide sequences of goose and muscovy  
RT duck parvoviruses indicates common ancestral origin with adeno-  
RT associated virus 2.";  
RL Virology 212:562-573(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VIRULENT B;  
RA Zadori Z.;  
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U25749; AAA83231.1; -.  
DR InterPro; IPR001403; Parvo\_coat.  
DR Pfam; PF00740; Parvo\_coat; 1.  
SQ SEQUENCE 587 AA; 65246 MW; 0278E5D5FE7F0423 CRC64;

Query Match 58.4%; Score 1698; DB 12; Length 587;  
Best Local Similarity 57.1%; Pred. No. 6.3e-117;  
Matches 310; Conservative 87; Mismatches 128; Indels 18; Gaps 8;

QY 1 MASGGAPMADNNEGADGVGNASGNWHCDSTWLGDRTVITSTRTWALPTYNNHLYKQISS 60  
DB 54 MAEGGGAGMGDSSSGADGVGNASGNWHCDSQMGNTVITKTRTWLPSYNNHIYKAITS 113  
QY 61 ASTGASNDNH--YFGYSTPWGYFDENRFHCHFSPRDMQRLINNNWGFRPKRLNFKLFNIQ 118  
DB 114 ---GTSQDANVOYAGYSTPWGYFDENRFHCHFSPRDMQRLINNNWGIRPKSLKFKIFNVQ 170  
QY 119 VKEVTTNDGVTTIANNLTSTVQVFSDEYQLPYVLGSAHQGLPPPADVFMIPOYGYLT 178

Db 171 VKEVTTODQTKTIANNLSTIQVFTDDEHQLPYVLGSATEGTPPFPSPDVYALPQYGYCT 230  
Qy 179 LN--NGSOAVGRSSFYCLEYFPSSOMLRTGNNTFSYTEEBV?FHSYAHQSOLDRLMNP 235  
Db 231 MHTNONGARFNDRSAFYCLEYFPSSOMLRTGNNTFSYTEEBV?FHSYAHQSOLDRLMNP 290  
Qy 236 LIDQYLYLNRTQNGSSAQNKDLLFSRGSPPAGMSVQPKNMLPGPCYRQORV-SKTXTDN 294  
Db 291 LVDQYLMNFNEV-DSSRNAQ-----FKKAVKAGYGTMRNMLPGPKFLDQVRAYTGGTD 344  
Qy 295 NNSNFT-WTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMGVMIFGKE--SAGASNT 351  
Db 345 NYANMNIMWSNGNKVNLKDRQYLLQPGPVSATYTEGEASSLPQNTLIGIAKDPYRSGSTTA 404  
Qy 352 ALDNVMTDEEIKATNPVATERFGTVAVNFQSSSTDPAIGDVHAMGALPGMWQDRDXY 411  
Db 405 GISDIMVTEEQEVAPTINGVWKPYGRTVTNEQNTTAPTSSDLVDLGLPGMWQNRDIY 464  
Qy 412 LQGPIMAKIPHTDGHFHPSPMLMGFGGLKNPPQILIKNTVPANPPAEFSATKFSFITQ 471  
Db 465 LQGPIGAKIPKTDGKHPSPNLGGFGHLNPPQVFIKNTVPADPPVEYVHOKMNSYITQ 524  
Qy 472 YSTGOVSVEIEWELQKENSQRWNPVOYTSNYAKSANDFTVDNNGLYTEPRPIGTRYLT 531  
Db 525 YSTGOCTVEMWELRKENSQRWNPVOYTSNYAKSANDFTVDNNGLYTEPRPIGTRYLT 584  
Qy 532 RPL 534  
Db 585 QNL 587

RESULT 10

ID 067666 PRELIMINARY; PRT; 732 AA.

AC 067666; 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE VP1.  
GN VP1.  
OS goose parvovirus.  
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.  
OX NCBI\_TaxID=38251;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VIRULENT B;  
RX MEDLINE=96010229; PubMed=7571426;  
RA Zadori Z., Stefancsik R., Rauch T., Kisary J.;  
RT "Analysis of the complete nucleotide sequences of goose and muscovy  
RT duck parvoviruses indicates common ancestral origin with adeno-  
RT associated virus 2."  
RL Virology 212:562-573 (1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VIRULENT B;  
RA Zadori Z.;  
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U25749; AAA83230.1; -.  
DR InterPro; IPR001403; Parvo\_coat.  
DR Pfam; PF00740; Parvo\_coat; 1.  
SQ SEQUENCE 732 AA; 81340 MW; 339507C61D47B52C CRC64;

Query Match 58.4%; Score 1698; DB 12; Length 732;  
Best Local Similarity 57.1%; Pred. No. 8.6e-117;  
Matches 310; Conservative 87; Mismatches 128; Indels 18; Gaps 8;

Qy 1 MASGGGAPMADNNEGADGVGNASGNWCHDSTWLGDRVITTSRTWTALPTYNNHLYKQISS 60  
Db 199 MAEGGGAGMGDSGGADGVGNASGNWCHDSDQMGNTVITKTRTWLPSYNNHIYKAITS 258  
Qy 61 ASTGASNDNH--YFGYSTPWGYFDNFRFCHFSPRDWQRLINNNWGFPRKRLNFKLFNIQ 118  
Db 259 ---GTSQDANVQYAGYSTPWGYFDNFRFCHFSPRDWQRLINNNWGIKPKSLKFIENVO 315

Qy 119 VKEVTNDGVTTIANNLSTVQVSDSEYQLPYVLGSAHQGLPFPADVEMIPQYGYLT 178  
Db 316 VKEVTTODQTKTIANNLSTIQVFTDDEHQLPYVLGSATEGTPPFPSPDVYALPQYGYCT 375  
Qy 179 LN--NGSOAVGRSSFYCLEYFPSSOMLRTGNNTFSYTEEBV?FHSYAHQSOLDRLMNP 235  
Db 376 MHTNONGARFNDRSAFYCLEYFPSSOMLRTGNNTFSYTEEBV?FHSYAHQSOLDRLMNP 435  
Qy 236 LIDQYLYLNRTQNGSSAQNKDLLFSRGSPPAGMSVQPKNMLPGPCYRQORV-SKTXTDN 294  
Db 436 LVDQYLMNFNEV-DSSRNAQ-----FKKAVKAGYGTMRNMLPGPKFLDQVRAYTGGTD 489  
Qy 295 NNSNFT-WTGASKYNLNGRESIINPGTAMASHKDEDEKFFPMGVMIFGKE--SAGASNT 351  
Db 490 NYANMNIMWSNGNKVNLKDRQYLLQPGPVSATYTEGEASSLPQNTLIGIAKDPYRSGSTTA 549  
Qy 352 ALDNVMTDEEIKATNPVATERFGTVAVNFQSSSTDPAIGDVHAMGALPGMWQDRDXY 411  
Db 550 GISDIMVTEEQEVAPTINGVWKPYGRTVTNEQNTTAPTSSDLVDLGLPGMWQNRDIY 609  
Qy 412 LQGPIMAKIPHTDGHFHPSPMLMGFGGLKNPPQILIKNTVPANPPAEFSATKFSFITQ 471  
Db 610 LQGPIGAKIPKTDGKHPSPNLGGFGHLNPPQVFIKNTVPADPPVEYVHOKMNSYITQ 669  
Qy 472 YSTGOVSVEIEWELQKENSQRWNPVOYTSNYAKSANDFTVDNNGLYTEPRPIGTRYLT 531  
Db 670 YSTGOCTVEMWELRKENSQRWNPVOYTSNYAKSANDFTVDNNGLYTEPRPIGTRYLT 729  
Qy 532 RPL 534  
Db 730 QNL 732

RESULT 11

ID 08V395 PRELIMINARY; PRT; 732 AA.

AC 08V395; 01-MAR-2002 (Tremblrel. 20, Created)  
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)  
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
DE Capsid protein VP.  
GN VP.  
OS goose parvovirus.  
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.  
OX NCBI\_TaxID=38251;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=GPV-YG;  
RA Ge Y., You Y., Xu Q.;  
RT "Analysis of the major open reading frames' nucleotide sequences in  
RT Goose parvovirus GPV-YG strain isolated in China."  
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF416726; ALJ37722.1; -.  
DR InterPro; IPR001403; Parvo\_coat.  
DR Pfam; PF00740; Parvo\_coat; 1.  
SQ SEQUENCE 732 AA; 81456 MW; 73F2E4BC769744B6 CRC64;

Query Match 58.4%; Score 1697; DB 12; Length 732;  
Best Local Similarity 57.3%; Pred. No. 1e-116;  
Matches 311; Conservative 82; Mismatches 132; Indels 18; Gaps 8;

Qy 1 MASGGGAPMADNNEGADGVGNASGNWCHDSTWLGDRVITTSRTWTALPTYNNHLYKQISS 60  
Db 199 MAEGGGAGMGDSGGADGVGNASGNWCHDSDQMGNTVITKTRTWLPSYNNHIYKAITS 258  
Qy 61 ASTGASNDN--HYFGYSTPWGYFDNFRFCHFSPRDWQRLINNNWGFPRKRLNFKLFNIQ 118  
Db 259 ---GTSQDATVQYAGYSTPWGYFDNFRFCHFSPRDWQRLINNNWGIKPKSLKFIENVO 315  
Qy 119 VKEVTNDGVTTIANNLSTVQVSDSEYQLPYVLGSAHQGLPFPADVEMIPQYGYLT 178  
Db 316 VKEVTTODQTKTIANNLSTIQVFTDDEHQLPYVLGSATEGTPPFPSPDVYALPQYGYCT 375

OY	179	LN---NGSQAVGRSSFFCYCLEYFPSPOMLRTGNPFBSYTFEEVFPFHSSYAHQSQSLDRIMNP	235
Dd	376	MHTNONGARFNDRSAPFYCLEYFPSOMLRTGNMFEFTDFEEVPFHSMAFSQDLDRIMNP	435
Oy	236	LIDOLYYLNRTONOGSGSAQNKDLLFSRGSFAGNSVO'PKMWLPGPCYRQRV-SKTXTDN	294
Dd	436	LYDQYLMMNFNEV-DSSRKAQ----FKKAVKGAYGTMRNMWLPGPCLLDQVRVRYTGGTD	489
Oy	295	NNSNFT-WTGASKYNLNGRESIIINPGTAMASHKHODEDKCFPMGSMIFGKE--SAGASNT	351
Dd	490	NYANMNINWSNGNKVNLLKDRO'YLLQPGPVSATHTKYEASSIPAQNILGLAKDPYRSGSTTA	549
Oy	352	ALDNVMITDEEBEIKATNPVATERFGTVAVNFQSSSTDPTATGDVHAMGALPGMWQDRDVY	411
Dd	550	GISDIMVTDEQEVAFTNGVGWKPYGKTYTNENQTTPAPTSDDLVLGALPGMWQNRDIY	609
Oy	412	LOGPIWAKIPHTDGHFHPSPLMGGFGLKNPPDILIKNTPVPANPPAEFSATKFASFITO	471
Dd	610	LOGPIWAKIPKTDGKFHPSPNLGFGFLYNPEPOVFIKNTPVPADPPEVEYVHQKMNSYITQ	669
Oy	472	YSTGOVSVEIEMBLQOKENSKRMNPEVOYTSNYAKSANVDFTVDNNGLYTEPRPLIGTRYLT	531
Dd	670	YSSGOCTVEMVWEHLRKENSKRWNPEIQFTSNFSDRTSIMFAPNETGGYIEDRLIGTRYLT	729
Oy	532	RPL 534	
Dd	730	QNL 732	

## RESULT 12

ID	041855	PRELIMINARY;	PRT;	734 AA.
DT	041855;			
DC	01-JAN-1998 (TREMBLrel. 05, Created)			
DT	01-JAN-1998 (TREMBLrel. 05, last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, last annotation update)			
DE	Capsid.			
OS	adeno-associated virus 4.			
OC	viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.			
OX	NCBI_TaxID=57579;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=ATCC VR-646;			
RX	MEDLINE=97404695; PubMed=9261407;			
RA	Chlorini J.A., Yang L., Liu Y., Safer B., Kotin R.M.;			
RT	"Cloning of adeno-associated virus type 4 (AAV4) and generation of			
RL	recombinant AAV4 particles.";			
RL	J. Virol. 71:6823-6833(1997).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=ATCC VR-646;			
RA	Chlorini J.A., Yang L., Kotin R.M., Safer B.;			
RL	Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: U89790; AAC58045.1; -.			
DR	InterPro; IPR001403; Parvo_coat.			
DR	Pfam; PF00740; Parvo_coat_1.			
SO	SEQUENCE 734 AA; 80639 MW; 616CC27A777BBE6F CRC64;			

Query Match	58.2%	Score 1690.5;	DB 12;	Length 734;
Best Local Similarity	59.4%;	Pred. No. 3.1e-116;		
Matches 325; Conservative	65;	Mismatches 132;	Indels 25;	Gaps 9;

[illegible]

Db	314	VTTSNGETTVANNLTSTVQIFPADSSYEIPYMDAGQEGSLPPFPNDVEMVPOYGCGLV	373
QY	179	LNNGSQAVGRSSFYCLEYFPQMLRTGNNFTFSYTFEEVPFHSSYAHQSIDLRLMPLID	238
Db	374	GNTSQOQTDNRNAFYCLEYFPQMLRTGNNFEITYSFEKVPFHSWYAHQSIDLRLMPLID	433
QY	239	QYLYLNRTON---QSGSAQNKDILFSRGSFAGMSVQPKNWLPGPCYRQQRVSKTKTDN	294
Db	434	QYLMGLQSTTTGTTLNAGTATN---FTKLRTPTNFSNFKKNWLPGPSTIKQGFSKTA--N	488
QY	295	NNSNFTWTGAS---KY---NLNGRESIINPGTAMASHKODEDKFFPMGVMIFGKESAG	347
Db	489	QNYKIPATGSDSLIKYETHSTLDGRWSALTTPGPMATAGPADSK-FSNSQLIFAGPKONG	547
QY	348	ASNTALDNVMTDEEEIKATNPVATERFGTAVNVFQSSSTDPAIGDVHAMGALPGMWOD	407
Db	548	NTATVPPTLI FTSEELATNA TDTDMG NLP GGDQSN SNLP TVDRLTALGAVPGMWQN	607
QY	408	RDVYLQGPWAKIPHTDGHFHSPLMGFGGLKNPPQILIKNTPVANPAPAEFSATKFA	467
Db	608	RDIIYQGPWAKIPHTDGHFHSPLIGFGGLKHPPOIFIKNTPVANPATTFSSSTPVNS	667
QY	468	FITQYSTGVSVIEIEMELQKENS KRNPNPEVQYTSNYAKSANDFTVDNNGLYTEPRPIGT	527
Db	668	FITQYSTGVSVQIDWEIQKENS KRNPNPEVQYTSNYGQONS LMAPDAGKYTEPRAIGT	727
QY	528	RYLTRPL 534	
Db	728	RYLTRHL 734	

## RESULT 13

ID	Q83290	PRELIMINARY;	PRT;	732 AA.
AC	Q83290;			
DT	01-NOV-1996 (TREMBLrel. 01, Created)			
DT	01-NOV-1996 (TREMBLrel. 01, last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, last annotation update)			
DE	Capsid protein.			
OS	Muscovy duck parvovirus.			
OC	Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.			
OX	NCBI_TaxID=37325;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=89384/France;			
RX	MEDLINE=96406928; PubMed=8811015;			
RA	Le Gall-Recule G., Jestin V., Chagnaud P., Blanchard P., Jestin A.;			
RT	"Expression of muscovy duck parvovirus capsid proteins (VP2 and VP3)			
RT	in a baculovirus expression system and demonstration of immunity			
RT	induced by the recombinant protein.";			
RL	J. Gen. Virol. 77:2159-2163(1996).			
DR	EMBL; Z68272; CAA92575.1; -			
DR	InterPro; IPR001403; Parvo_coat.			
DR	Pfam; PF00740; Parvo_coat; 1.			
FT	CHAIN 146 732			
FT	VP2 CAPSID PROTEIN.			
FT	CHAIN 199 732			
FT	VP3 CAPSID PROTEIN.			
SEQUENCE	732 AA; 81364 MW; DE70DCCAB215F4E2 CRC64;			

Query Match	58.2%;	Score 1690;	DB 12;	Length 732;
Best Local Similarity	57.3%;	Pred. No. 3.3e-116;		
Matches 310;	Conservative 76;	Mismatches 141;	Indels 14;	Gaps 6;

```
QY      1 MASGGGAPMADNNEGADGVGNASGNWHCHDSTWLGDRIITSTRTIALPTYYNNHLKYKQISS    60
        ||| : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     199 MAEGSGAMGDSAGAGADVGNASGNWHCHDSQWLGDTVITKTTRTWLPSPYNNHIYKAITS    258

QY      61 ASTGASNNDNHFGYSTPMGWEDFNRFHCHFSPRDWQRLLNNNWGFEPKRLNFKLFINIQVK    120
        : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     259 GTNPDSN-TQYAGYSTPMWGTFDNFRFHCHFSPRDWQRLLNNHWGRIPKALKFKIFNVQVK    317

QY      121 EYTTNDGVTIIANNLTSTVOVFSDSEYOQLPYVLGSAHQGCLPEPPADVFMIPOYGYLTLN    180
        |||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     318 EVTTQDOTKIANNLTSTIOIFTDNEHQPLVYLSATEGTMPPEPSDVYALPOYGCTMH    377
```







